



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C07K 14/00</b>	<b>A2</b>	(11) International Publication Number: <b>WO 99/33871</b>
		(43) International Publication Date: <b>8 July 1999 (08.07.99)</b>
<p>(21) International Application Number: <b>PCT/US98/27918</b></p> <p>(22) International Filing Date: <b>30 December 1998 (30.12.98)</b></p> <p>(30) Priority Data: <b>60/070,116</b>      <b>31 December 1997 (31.12.97)</b>      <b>US</b></p> <p>(71) Applicant: <b>MILLENNIUM PHARMACEUTICALS, INC.</b> [US/US]; 640 Memorial Drive, Cambridge, MA 02139 (US).</p> <p>(72) Inventors: <b>YOUNGMAN, Philip</b>; 92 Charles Street #53, Boston, MA 02114 (US). <b>FRITZ, Christian</b>; 14 BROADS Avenue, Natick, MA 01760 (US). <b>MURPHY, Christopher</b>; 7 Warren Street, Upton, MA 01568 (US). <b>GUZMAN, Luz-Maria</b>; 52 Athol Street, Boston, MA 02134 (US).</p> <p>(74) Agents: <b>FASSE, J., Peter et al</b>; Fish &amp; Richardson P.C., 225 Franklin Street, Boston, MA 02110-2804 (US).</p>		<p>(81) Designated States: <b>AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</b></p> <p><b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i></p>
<p>(54) Title: <b>ESSENTIAL BACTERIAL GENES AND THEIR USE</b></p> <p>(SEQ ID NO: 2) <sup>1</sup> <b>TGCTGATTTTGGGAAAGTTTATAGAGTTAAAGAGCTTAAGGAAAAAATTCATTGATATTTCTCTATATAAATAGATAAAATGCTACAAATG</b> 100 <b>AGACTAAAAAGCTCTTTCAATTAAGCTCTATTTTCTGAGATTCCTTTTAAAGTTAACTTAAGAAAGAAATATTTATCATTTTGAAGTTTAT</b></p> <p>(SEQ ID NO: 3)</p> <p>101 <b>ATAAATGAGGTAATAGGATAGGTTAGATAAATATTTAAAGATCTCCGATTTATCAAGGCTCCTACGCTCCCAAGCAATGTCAGATAAAGTAA</b> 200 <b>TATTGAAGTCAATATTCCTACTCTAATCTATTATAAATTTTCTAGAGCTTAAAGTTCCGAGCTGCTGAGCTTTCTCTCATCTCTATTTCCTCT</b></p> <p>(SEQ ID NO: 1) <sup>1</sup> <b>M R L D E Y L E V S I I E E T V A K E V A D E R</b> 27</p> <p>201 <b>ATCAAGGTTAAAGGATCTTGGCCAAAGTTCAAGGCTTGAAGTTAATGACCAAGTTGAAGTTCCCTTTGCAATAAGTTCTCTCTTGTAAAGTAC</b> 300 <b>TAACTGCAATTACCTTAGAGAGGCTTTTCAAGTTGCTGCAAGTTTCAATTAAGTTTCACTTGAAGGAAAGCTTATTCAGAGAGGAGCTTTTGTG</b></p> <p>30 <b>I E V N G I L A E S T D L E V D O V E I A P G N L L L V E V L</b> 61</p> <p>301 <b>TAGAGTGAAGATATGACAAAAAGAGATGACAGAGAAATGTAAGTTTCCAGTAAAGCTGCTGAGAGAAATGCTTAAGAAATATTTGACAT</b> 400 <b>ATCTCTACTTTCTATCAATTTTCTCTAGAGCTGCTTACATCTTTAATAGTCACTTTGTGAGGATCTCTTTTACAGATTTTATACATGTTA</b></p> <p>62 <b>S N R D S Y E E D A A G N Y I I E S T R V E E V</b> 89</p>		
<p>(57) Abstract</p> <p>Disclosed are 23 genes, termed "GEP" genes, found in <i>streptococcus pneumonia</i>, which are located within operons that are essential for survival. Also disclosed is a related essential gene found in <i>Bacillus subtilis</i>. These genes and the polypeptides that they encode, as well as homologs thereof, can be used to identify antibacterial agents for treating bacterial infections such as streptococcal pneumonia.</p>		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon		Republic of Korea	PT	Portugal		
CN	China	KR	Republic of Korea	RO	Romania		
CU	Cuba	KZ	Kazakhstan	RU	Russian Federation		
CZ	Czech Republic	LC	Saint Lucia	SD	Sudan		
DE	Germany	LI	Liechtenstein	SE	Sweden		
DK	Denmark	LK	Sri Lanka	SG	Singapore		
EE	Estonia	LR	Liberia				

- 1 -

ESSENTIAL BACTERIAL GENES AND THEIR USEBackground of the Invention

The invention relates to essential bacterial genes and their use in identifying antibacterial agents.

5        Bacterial infections may be cutaneous, subcutaneous, or systemic. Opportunistic bacterial infections proliferate, especially in patients afflicted with AIDS or other diseases that compromise the immune system. The bacterium *Streptococcus pneumonia* typically infects the respiratory tract and can cause lobar pneumonia, as well as meningitis, sinusitis, and other infections.

10        Summary of the Invention

The invention is based on the discovery of 23 genes in the bacterium *Streptococcus pneumoniae*, and a related gene in the bacterium *Bacillus subtilis*, that are located within operons that are essential for survival. These 23 *Streptococcus* genes are referred to herein as "GEP genes" (which stands for  
15    general essential protein); for convenience, the polypeptides encoded by these genes are referred to herein as "GEP polypeptides." Each GEP gene is located within an operon that contains a gene that is essential for survival of *Streptococcus pneumoniae*; the essential gene can be the GEP gene or another gene located within the same operon. Bacterial operons contain several genes that are related, e.g.,  
20    with respect to function or biochemical pathway. Transcription of an operon leads to the production of a single transcript in which multiple coding regions are linked. Thus, an operon containing one or more essential genes can be considered an "essential operon," since disruption of expression of one gene located within the operon will interfere with expression of the other genes in the operon. Each coding  
25    region of the transcript is separately translated into an individual polypeptide by ribosomes that initiate translation at multiple points along the transcript. Having identified one gene in the operon, one can readily identify and sequence the other genes located within the operon.

- 2 -

The genes encoding the GEP polypeptides are useful molecular tools for identifying similar genes in pathogenic microorganisms, such as pathogenic strains of *Bacillus*. In addition, the operons containing genes encoding GEP polypeptides, and the polypeptides encoded by such operons, are useful targets for identifying  
5 compounds that are inhibitors of the pathogens in which the GEP polypeptides are expressed. Such inhibitors inhibit bacterial growth by being bacteriostatic (e.g., inhibiting reproduction or cell division) or by being bacteriocidal (i.e., by causing cell death).

The invention, therefore, features an isolated polypeptide encoded by a  
10 nucleic acid located within an operon encoding a GEP polypeptide, termed gep103, having the amino acid sequence set forth in SEQ ID NO:1, or conservative variations thereof. An isolated operon comprising a nucleic acid encoding gep103 also is included within the invention. In addition, the invention includes an isolated nucleic acid of (a) an operon comprising the sequence of SEQ ID NO:2, as  
15 depicted in Fig. 1, or degenerate variants thereof; (b) an operon comprising the sequence of SEQ ID NO:2, or degenerate variants thereof, wherein T is replaced by U; (c) nucleic acids complementary to (a) and (b); and (d) fragments of (a), (b), and (c) that are at least 15 base pairs in length and that hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:1. As  
20 described above for gep103, other nucleic acids and polypeptides encoded by nucleic acids located within operons encoding GEP polypeptides are included within the invention, including: (a) operons comprising the nucleic acids represented by the SEQ ID NOs. listed below, as depicted in the Figures listed below, or degenerate variants thereof; (b) operons comprising the nucleic acids  
25 represented by the SEQ ID NOs. listed below, wherein T is replaced by U; (c) nucleic acids complementary to (a) and (b); and (d) fragments of (a), (b), and (c) that are at least 15 base pairs in length and that hybridize under stringent conditions to genomic DNA encoding the polypeptides represented by the SEQ ID NOs. listed below.



- 3 -

Table 1: GEP nucleic acids and polypeptides

	GEP Nucleic Acid or Polypeptide	Figure No.	SEQ ID No. of Amino Acid Sequence	SEQ ID No. of the Coding Strand of the Nucleic Acid Sequence	SEQ ID No. of the Non-coding Strand of the Nucleic Acid Sequence
5	gep103	1	1	2	3
	gep1119	2	4	5	6
	gep1122	3	7	8	9
	gep1315	4	10	11	12
10	gep1493	5	13	14	15
	gep1507	6	16	17	18
	gep1511	7	19	20	21
	gep1518	8	22	23	24
15	gep1546	9	25	26	27
	gep1551	10	28	29	30
	gep1561	11	31	32	33
	gep1580	12	34	35	36
20	gep1713	13	37	38	39
	gep222	14	40	41	42
	gep2283	15	43	44	45
	gep273	16	46	47	48
25	gep286	17	49	50	51
	gep311	18	52	53	54
	gep3262	19	55	56	57
	gep3387	20	58	59	60
	gep47	21	61	62	63

- 4 -

GEP Nucleic Acid or Polypeptide	Figure No.	SEQ ID No. of Amino Acid Sequence	SEQ ID No. of the Coding Strand of the Nucleic Acid Sequence	SEQ ID No. of the Non-coding Strand of the Nucleic Acid Sequence
gep61	22	64	65	66
gep76	23	67	68	69

The invention also includes allelic variants (i.e., genes encoding isozymes) of the genes located within operons encoding the GEP polypeptides listed above.

- 5 For example, the invention includes a gene that encodes a GEP polypeptide but which gene includes one or more point mutations, deletions, promotor variants, or splice site variants, provided that the resulting GEP polypeptide functions as a GEP polypeptide (e.g., as determined in a conventional complementation assay).

- Identification of these GEP genes and the determination that they are
- 10 located within operons containing an essential gene allows homologs of the GEP genes to be found in other organisms strains of *Streptococcus*. Also, orthologs of these genes can be identified in other species (e.g., *Bacillus sp.*). While "homologs" are structurally similar genes contained within a species, "orthologs" are functionally equivalent genes from other species (within or outside of a given
- 15 genus, e.g., from *Bacillus subtilis* or *E. coli*). Such homologs and orthologs are expected to be located within operons that are essential for survival. Such homologous and orthologous genes and polypeptides can be used to identify compounds that inhibit the growth of the host organism (e.g., compounds that are bacteriocidal or bacteriostatic against pathogenic strains of the organism).
- 20 Homologous and orthologous genes and polypeptides that are essential for survival can serve as targets for identifying a broad spectrum of antibacterial agents.

An ortholog of gep1493, termed B-yneS, has been identified in *B. subtilis* and is essential for survival of *B. subtilis*. The amino acid sequence (SEQ ID NO: 70), coding sequence (SEQ ID NO:71), and non-coding sequence (SEQ ID NO:72)

- 5 -

of B-yneS is set forth in Fig. 24. As with the other polypeptides and genes disclosed herein, the B-yneS polypeptide and gene can be used in the methods described herein to identify antibacterial agents.

The term gep103 polypeptide or gene as used herein is intended to include  
5 the polypeptide and gene set forth in Fig. 1 herein, as well as homologs of the sequences set forth in Fig. 1. Also encompassed by the term gep103 gene are degenerate variants of the nucleic acid sequence set forth in Fig. 1 (SEQ ID NO:2). Degenerate variants of a nucleic acid sequence exist because of the degeneracy of the amino acid code; thus, those sequences that vary from the sequence represented  
10 by SEQ ID NO:2, but which nonetheless encode a gep103 polypeptide are included within the invention. Likewise, because of the similarity in the structures of amino acids, conservative variations (as described herein) can be made in the amino acid sequence of the gep103 polypeptide while retaining the function of the polypeptide (e.g., as determined in a conventional complementation assay). Other gep103  
15 polypeptides and genes identified in additional *Streptococcus* strains may be such conservative variations or degenerate variants of the particular gep103 polypeptide and nucleic acid set forth in Fig. 1 (SEQ ID NOs:1 and 2, respectively). The gep103 polypeptide and gene share at least 80%, e.g., 90%, sequence identity with SEQ ID NOs:1 and 2, respectively. Regardless of the percent sequence identity  
20 between the gep103 sequence and the sequence represented by SEQ ID NOs:1 and 2, the gep103 genes and polypeptides encompassed by the invention are able to complement for the lack of gep103 function (e.g., in a temperature-sensitive mutant) in a standard complementation assay. Additional gep103 genes that are identified and cloned from additional *Streptococcus* strains, and pathogenic strains  
25 in particular, can be used to produce gep103 polypeptides for use in the various methods described herein, e.g., for identifying antibacterial agents. Likewise, the terms gep1119, gep1122, gep1315, gep1493, gep1507, gep1511, gep1518, gep1546, gep1551, gep1561, gep1580, gep1713, gep222, gep2283, gep273, gep286, gep311, gep3262, gep3387, gep47, gep61, and gep76 encompass homologs, conservative  
30 variations, and degenerate variants of the sequences depicted in Figs. 2-23,

- 6 -

respectively. Such homologs, conservative variations, and degenerate variants also are included within the invention.

Since the various GEP genes described herein have been identified and shown to be located within operons that are essential for survival, the GEP genes  
5 and polypeptides encoded by nucleic acid sequences located within operons containing GEP genes and their homologs and orthologs can be used to identify antibacterial agents. More specifically, the polypeptides encoded by nucleic acid sequences located within operons containing GEP genes can be used, separately or together, in assays to identify test compounds that bind to these polypeptides. Such  
10 test compounds are expected to be antibacterial agents, in contrast to compounds that do not bind to these GEP polypeptides. As described herein, any of a variety of art-known methods can be used to assay for binding of test compounds to the polypeptides. The invention includes, for example, a method for identifying an antibacterial agent where the method entails: (a) contacting a polypeptide encoded  
15 by a nucleic acid sequence located within an operon containing a GEP gene, or homolog or ortholog thereof, with a test compound; (b) detecting binding of the test compound to the polypeptide or homolog or ortholog; and (c) determining whether a test compound that binds to the polypeptide or homolog or ortholog inhibits growth of bacteria, relative to growth of bacteria cultured in the absence of  
20 the test compound that binds to the polypeptide or homolog or ortholog, as an indication that the test compound is an antibacterial agent.

In various embodiments, the GEP polypeptide is derived from a non-pathogenic or pathogenic *Streptococcus* strain, such as *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Streptococcus agalactiae*, *Streptococcus endocarditis*,  
25 *Streptococcus faecium*, *Streptococcus sanguis*, *Streptococcus viridans*, and *Streptococcus hemolyticus*. Suitable orthologs of the *Streptococcus* GEP genes can be derived from the bacterium *Bacillus subtilis*. The test compound can be immobilized on a substrate, and binding of the test compound to the polypeptide or homolog or ortholog can be detected as immobilization of the polypeptide or

- 7 -

homolog or ortholog on the immobilized test compound, e.g., in an immunoassay with an antibody that specifically binds to the polypeptide.

If desired, the test compound can be a test polypeptide (e.g., a polypeptide having a random or predetermined amino acid sequence; or a naturally-occurring or synthetic polypeptide). Alternatively, the test compound can be a nucleic acid, such as a DNA or RNA molecule. In addition, small organic molecules can be tested. The test compound can be a naturally-occurring compound or it can be synthetically produced, if desired. Synthetic libraries, chemical libraries, and the like can be screened to identify compounds that bind to the polypeptides. More generally, binding of test compounds to the polypeptide or homolog or ortholog can be detected either *in vitro* or *in vivo*. Regardless of the source of the test compound, the polypeptides described herein can be used to identify compounds that are bacterioid or bacteriostatic to a variety of pathogenic or non-pathogenic strains.

15 In an exemplary method, binding of a test compound to a polypeptide encoded by a nucleic acid located within an operon containing a GEP gene can be detected in a conventional two-hybrid system for detecting protein/protein interactions (e.g., in yeast or mammalian cells). Generally, in such a method, (a) the polypeptide encoded by a nucleic acid located within an operon containing a GEP gene is provided as a fusion protein that includes the polypeptide fused to (i) a transcription activation domain of a transcription factor or (ii) a DNA-binding domain of a transcription factor; (b) the test polypeptide is provided as a fusion protein that includes the test polypeptide fused to (i) a transcription activation domain of a transcription factor or (ii) a DNA-binding domain of a transcription factor; and (c) binding of the test polypeptide to the polypeptide is detected as reconstitution of a transcription factor. Homologs and orthologs of the GEP polypeptides can be used in similar methods. Reconstitution of the transcription factor can be detected, for example, by detecting transcription of a gene that is operably linked to a DNA sequence bound by the DNA-binding domain of the reconstituted transcription factor (See, for example, White, 1996, Proc. Natl. Acad.

20  
25  
30

- 8 -

Sci. 93:10001-10003 and references cited therein and Vidal et al., 1996, Proc. Natl. Acad. Sci. 93:10315-10320).

In an alternative method, an isolated operon containing a nucleic acid molecule encoding a GEP polypeptide is used to identify a compound that  
5 decreases the expression of a GEP polypeptide *in vivo*. Such compounds can be used as antibacterial agents. To discover such compounds, cells that express a GEP polypeptide are cultured, exposed to a test compound (or a mixture of test compounds), and the level of expression or activity is compared with the level of GEP polypeptide expression or activity in cells that are otherwise identical but that  
10 have not been exposed to the test compound(s). Many standard quantitative assays of gene expression can be utilized in this aspect of the invention.

To identify compounds that modulate expression of a GEP polypeptide (or homologous or orthologous sequence), the test compound(s) can be added at varying concentrations to the culture medium of cells that express a GEP  
15 polypeptide (or homolog or ortholog), as described herein. Such test compounds can include small molecules (typically, non-protein, non-polysaccharide chemical entities), polypeptides, and nucleic acids. The expression of the GEP polypeptide is then measured, for example, by Northern blot PCR analysis or RNase protection analyses using a nucleic acid molecule of the invention as a probe. The level of  
20 expression in the presence of the test molecule, compared with the level of expression in its absence, will indicate whether or not the test molecule alters the expression of the GEP polypeptide. Because the GEP polypeptides are expressed from operons that are essential for survival, test compounds that inhibit the expression and/or function of the GEP polypeptide will inhibit growth of the cells  
25 or kill the cells.

Compounds that modulate the expression of the polypeptides of the invention can be identified by carrying out the assays described herein and then measuring the levels of the GEP polypeptides expressed in the cells, e.g., by performing a Western blot analysis using antibodies that bind to a GEP  
30 polypeptide.

- 9 -

- The invention further features methods of identifying from a large group of mutants those strains that have conditional lethal mutations. In general, the gene and corresponding gene product are subsequently identified, although the strains themselves can be used in screening or diagnostic assays. The mechanism(s) of action for the identified genes and gene products provide a rational basis for the design of antibacterial therapeutic agents. These antibacterial agents reduce the action of the gene product in a wild type strain, and therefore are useful in treating a subject with that type, or a similarly susceptible type of infection by administering the agent to the subject in a pharmaceutically effective amount.
- 5    Reduction in the action of the gene product includes competitive inhibition of the gene product for the active site of an enzyme or receptor; non-competitive inhibition; disrupting an intracellular cascade path which requires the gene product; binding to the gene product itself, before or after post-translational processing; and acting as a gene product mimetic, thereby down-regulating the activity.
- 10    Therapeutic agents include monoclonal antibodies raised against the gene product.

Furthermore, the presence of the gene sequence in certain cells (e.g., a pathogenic bacterium of the same genus or similar species), and the absence or divergence of the sequence in host cells can be determined, if desired. Therapeutic agents directed toward genes or gene products that are not present in the host have several advantages, including fewer side effects, and lower overall dosage.

15    several advantages, including fewer side effects, and lower overall dosage.

- The invention includes pharmaceutical formulations that include a pharmaceutically acceptable excipient and an antibacterial agent identified using the methods described herein. In particular, the invention includes pharmaceutical formulations that contain antibacterial agents that inhibit the growth of, or kill, pathogenic *Streptococcus* strains. Such pharmaceutical formulations can be used for treating a *Streptococcus* infection in an organism. Such a method entails administering to the organism a therapeutically effective amount of the pharmaceutical formulation. In particular, such pharmaceutical formulations can be used to treat streptococcal pneumonia in mammals such as humans and domesticated mammals (e.g., cows, pigs, dogs, and cats), and in plants. The
- 20    pathogenic *Streptococcus* strains. Such pharmaceutical formulations can be used for treating a *Streptococcus* infection in an organism. Such a method entails administering to the organism a therapeutically effective amount of the pharmaceutical formulation. In particular, such pharmaceutical formulations can be used to treat streptococcal pneumonia in mammals such as humans and domesticated mammals (e.g., cows, pigs, dogs, and cats), and in plants. The
- 25    pathogenic *Streptococcus* strains. Such pharmaceutical formulations can be used for treating a *Streptococcus* infection in an organism. Such a method entails administering to the organism a therapeutically effective amount of the pharmaceutical formulation. In particular, such pharmaceutical formulations can be used to treat streptococcal pneumonia in mammals such as humans and domesticated mammals (e.g., cows, pigs, dogs, and cats), and in plants. The
- 30    domesticated mammals (e.g., cows, pigs, dogs, and cats), and in plants. The

- 10 -

efficacy of such antibacterial agents in humans can be estimated in an animal model system well known to those of skill in the art (e.g., mouse and rabbit model systems).

Also included within the invention are polyclonal and monoclonal antibodies  
5 that specifically bind to the various GEP polypeptides described herein (e.g., gep103). Such antibodies can facilitate detection of GEP polypeptides in various *Streptococcus* strains. These antibodies also are useful for detecting binding of a test compound to GEP polypeptides (e.g., using the assays described herein). In addition, monoclonal antibodies that bind to GEP polypeptides are themselves  
10 adequate antibacterial agents when administered to a mammal, as such monoclonal antibodies are expected to impede one or more functions of GEP polypeptides.

As used herein, "nucleic acids" encompass both RNA and DNA, including genomic DNA and synthetic (e.g., chemically synthesized) DNA. The nucleic acid can be double-stranded or single-stranded. Where single-stranded, the nucleic acid  
15 may be a sense strand or an antisense strand. The nucleic acid may be synthesized using oligonucleotide analogs or derivatives (e.g., inosine or phosphorothioate nucleotides). Such oligonucleotides can be used, for example, to prepare nucleic acids that have altered base-pairing abilities or increased resistance to nucleases.

An "isolated nucleic acid" is a DNA or RNA that is not immediately  
20 contiguous with both of the coding sequences with which it is immediately contiguous (one on the 5' end and one on the 3' end) in the naturally occurring genome of the organism from which it is derived. Thus, in one embodiment, an isolated nucleic acid includes some or all of the 5' non-coding (e.g., promoter) sequences that are immediately contiguous to the coding sequence. The term  
25 therefore includes, for example, a recombinant DNA that is incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (e.g., a genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other sequences. It also includes a recombinant DNA that is part of  
30 a hybrid gene encoding an additional polypeptide sequence. The term "isolated"



- 11 -

can refer to a nucleic acid or polypeptide that is substantially free of cellular material, viral material, or culture medium (when produced by recombinant DNA techniques), or chemical precursors or other chemicals (when chemically synthesized). Moreover, an "isolated nucleic acid fragment" is a nucleic acid  
5 fragment that is not naturally occurring as a fragment and would not be found in the natural state. As used herein, the term "isolated nucleic acid molecule" includes an operon containing a contiguous cluster of linked sequences. "Isolated operons" are those operons that are not naturally occurring and which are not associated with the sequences by which they are normally surrounded in a bacterial genome.

10 A nucleic acid sequence that is "substantially identical" to a GEP nucleotide sequence is at least 80% (e.g., 85%) identical to the nucleotide sequence of the nucleic acid sequences represented by the SEQ ID NOs listed in Table 1, as depicted in Figs. 1-23. For purposes of comparison of nucleic acids, the length of the reference nucleic acid sequence will generally be at least 40 nucleotides, e.g., at  
15 least 60 nucleotides or more nucleotides. Sequence identity can be measured using sequence analysis software (e.g., Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705).

The GEP polypeptides useful in practicing the invention include, but are not  
20 limited to, recombinant polypeptides and natural polypeptides. Also useful in the invention are nucleic acid sequences that encode forms of GEP polypeptides in which naturally occurring amino acid sequences are altered or deleted. Preferred nucleic acids encode polypeptides that are soluble under normal physiological conditions. Also within the invention are nucleic acids encoding fusion proteins in  
25 which a portion of a GEP polypeptide is fused to an unrelated polypeptide (e.g., a marker polypeptide or a fusion partner) to create a fusion protein. For example, the polypeptide can be fused to a hexa-histidine tag to facilitate purification of bacterially expressed polypeptides, or to a hemagglutinin tag to facilitate purification of polypeptides expressed in eukaryotic cells. The invention also  
30 includes, for example, isolated polypeptides (and the nucleic acids that encode these

- 12 -

polypeptides) that include a first portion and a second portion; the first portion includes, e.g., a GEP polypeptide, and the second portion includes an immunoglobulin constant (Fc) region or a detectable marker.

The fusion partner can be, for example, a polypeptide which facilitates secretion, e.g., a secretory sequence. Such a fused polypeptide is typically referred to as a preprotein. The secretory sequence can be cleaved by the host cell to form the mature protein. Also within the invention are nucleic acids that encode a GEP polypeptide fused to a polypeptide sequence to produce an inactive preprotein. Preproteins can be converted into the active form of the protein by removal of the inactivating sequence.

The invention also includes nucleic acids that hybridize, e.g., under stringent hybridization conditions (as defined herein) to all or a portion of the nucleotide sequences represented by the SEQ ID NOs. listed in Table 1, or their complements. The hybridizing portion of the hybridizing nucleic acids is typically at least 15 (e.g., 20, 30, or 50) nucleotides in length. The hybridizing portion of the hybridizing nucleic acid is at least 80%, e.g., at least 95%, or at least 98%, identical to the sequence of a portion or all of a nucleic acid encoding a GEP polypeptide or its complement. Hybridizing nucleic acids of the type described herein can be used as a cloning probe, a primer (e.g., a PCR primer), or a diagnostic probe. Nucleic acids that hybridize to the nucleotide sequences represented by the SEQ ID NOs. listed in Table 1 are considered "antisense oligonucleotides." Also included within the invention are ribozymes that inhibit the function of operons containing the GEP genes of the invention, as determined, for example, in a complementation assay.

Also useful in the invention are various cells, e.g., transformed host cells, that contain a GEP nucleic acid described herein. A "transformed cell" is a cell into which (or into an ancestor of which) has been introduced, by means of recombinant DNA techniques, a nucleic acid encoding a GEP polypeptide. Both prokaryotic and eukaryotic cells are included, e.g., bacteria, *Streptococcus*, *Bacillus*, and the like.

- 13 -

Also useful in the invention are genetic constructs (e.g., vectors and plasmids) that include a nucleic acid of the invention which is operably linked to a transcription and/or translation sequence to enable expression, e.g., expression vectors. By "operably linked" is meant that a selected nucleic acid, e.g., a DNA  
5 molecule encoding a GEP polypeptide, is positioned adjacent to one or more sequence elements, e.g., a promoter, which directs transcription and/or translation of the sequence such that the sequence elements can control transcription and/or translation of the selected nucleic acid.

The invention also features purified or isolated polypeptides encoded by  
10 nucleic acids located within operons containing GEP genes, as listed in Table 1. As used herein, both "protein" and "polypeptide" mean any chain of amino acids, regardless of length or post-translational modification (e.g., glycosylation or phosphorylation). Thus, the terms gep103 polypeptide, gep1119 polypeptide, gep1122 polypeptide, gep1315 polypeptide, gep1493 polypeptide, gep1507  
15 polypeptide, gep1511 polypeptide, gep1518 polypeptide, gep1546 polypeptide, gep1551 polypeptide, gep1561 polypeptide, gep1580 polypeptide, gep1713 polypeptide, gep222 polypeptide, gep2283 polypeptide, gep273 polypeptide, gep286 polypeptide, gep311 polypeptide, gep3262 polypeptide, gep3387 polypeptide, gep47 polypeptide, gep61 polypeptide, and gep76 polypeptide include full-length,  
20 naturally occurring gep103, gep1119, gep1122, gep1315, gep1493, gep1507, gep1511, gep1518, gep1546, gep1551, gep1561, gep1580, gep1713, gep222, gep2283, gep273, gep286, gep311, gep3262, gep3387, gep47, gep61, and gep76 proteins, respectively, as well as recombinantly or synthetically produced polypeptides that correspond to the full-length, naturally occurring proteins, or to a  
25 portion of the naturally occurring or synthetic polypeptide.

A "purified" or "isolated" compound is a composition that is at least 60% by weight the compound of interest, e.g., a GEP polypeptide or antibody. Preferably the preparation is at least 75% (e.g., at least 90% or 99%) by weight the compound of interest. Purity can be measured by any appropriate standard method, e.g.,  
30 column chromatography, polyacrylamide gel electrophoresis, or HPLC analysis.

- 14 -

Preferred GEP polypeptides include a sequence substantially identical to all or a portion of a naturally occurring GEP polypeptide, e.g., including all or a portion of the sequences shown in Figs. 1-23. Polypeptides "substantially identical" to the GEP polypeptide sequences described herein have an amino acid sequence  
5 that is at least 80% (e.g., 85%, 90%, 95%, or 99%) identical to the amino acid sequence of the GEP polypeptides represented by the SEQ ID NOs. listed in Table 1. For purposes of comparison, the length of the reference GEP polypeptide sequence will generally be at least 16 amino acids, e.g., at least 20 or 25 amino acids.

10 In the case of polypeptide sequences that are less than 100% identical to a reference sequence, the non-identical positions are preferably, but not necessarily, conservative substitutions for the reference sequence. Conservative substitutions typically include substitutions within the following groups: glycine and alanine; valine, isoleucine, and leucine; aspartic acid and glutamic acid; asparagine and  
15 glutamine; serine and threonine; lysine and arginine; and phenylalanine and tyrosine.

Where a particular polypeptide is said to have a specific percent identity to a reference polypeptide of a defined length, the percent identity is relative to the reference polypeptide. Thus, a polypeptide that is 50% identical to a reference  
20 polypeptide that is 100 amino acids long can be a 50 amino acid polypeptide that is completely identical to a 50 amino acid long portion of the reference polypeptide. It also might be a 100 amino acid long polypeptide which is 50% identical to the reference polypeptide over its entire length. Of course, other polypeptides also will meet the same criteria.

25 The invention also features purified or isolated antibodies that specifically bind to a GEP polypeptide. By "specifically binds" is meant that an antibody recognizes and binds to a particular antigen, e.g., a GEP polypeptide, but does not substantially recognize and bind to other molecules in a sample, e.g., a biological sample that naturally includes a GEP polypeptide.

- 15 -

In another aspect, the invention features a method for detecting a GEP polypeptide in a sample. This method includes: obtaining a sample suspected of containing a GEP polypeptide; contacting the sample with an antibody that specifically binds to a GEP polypeptide under conditions that allow the formation  
5 of complexes of an antibody and the GEP polypeptide; and detecting the complexes, if any, as an indication of the presence of a GEP polypeptide in the sample.

Also encompassed by the invention is a method of obtaining a gene related to (i.e., a functional homolog or ortholog of) a GEP gene. Such a method entails  
10 obtaining a labeled probe that includes an isolated nucleic acid which encodes all or a portion of a GEP nucleic acid, or a homolog or ortholog thereof; screening a nucleic acid fragment library with the labeled probe under conditions that allow hybridization of the probe to nucleic acid fragments in the library, thereby forming nucleic acid duplexes; isolating labeled duplexes, if any; and preparing a full-length  
15 gene sequence from the nucleic acid fragments in any labeled duplex to obtain a gene related to the GEP gene.

The invention offers several advantages. For example, the methods for identifying antibacterial agents can be configured for high throughput screening of numerous candidate antibacterial agents.

20 Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described herein. All  
25 publications, patent applications, patents, and other references mentioned herein are incorporated herein by reference in their entirety. In the case of a conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative and are not intended to limit the scope of the invention, which is defined by the claims.

- 16 -

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

#### Brief Description of the Drawings

Fig. 1 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep103 polypeptide and gene from a *Streptococcus pneumonia* strain (SEQ ID NOs:1, 2, and 3 respectively).

Fig. 2 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep1119 polypeptide and gene from a *Streptococcus pneumonia* strain (SEQ ID NOs:4, 5 and 6, respectively).

Fig. 3 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep1122 polypeptide and gene from a *Streptococcus pneumonia* strain (SEQ ID NOs:7, 8, and 9, respectively).

Fig. 4 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep1315 polypeptide and gene from a *Streptococcus pneumonia* strain (SEQ ID NOs:10, 11, and 12, respectively).

Fig. 5 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep1493 polypeptide and gene from a *Streptococcus pneumonia* strain (SEQ ID NOs:13, 14, and 15, respectively).

Fig. 6 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep1507 polypeptide and gene from a *Streptococcus pneumonia* (SEQ ID NOs:16, 17, and 18, respectively).

- 17 -

Fig. 7 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep1511 polypeptide and gene from a *Streptococcus pneumonia* (SEQ ID NOs:19, 20, and 21, respectively).

Fig. 8 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep1518 polypeptide and gene from a *Streptococcus pneumonia* (SEQ ID NOs:22, 23, and 24, respectively).

Fig. 9 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep1546 polypeptide and gene from a *Streptococcus pneumonia* strain (SEQ ID NOs:25, 26, and 27, respectively).

Fig. 10 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep1551 polypeptide and gene from a *Streptococcus pneumonia* strain (SEQ ID NOs:28, 29, and 30, respectively).

Fig. 11 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep1561 polypeptide and gene from a *Streptococcus pneumonia* strain (SEQ ID NOs:31, 32, and 33, respectively).

Fig. 12 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep1580 polypeptide and gene from a *Streptococcus pneumonia* strain (SEQ ID NOs:34, 35, and 36, respectively).

Fig. 13 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep1713 polypeptide and gene from a *Streptococcus pneumonia* (SEQ ID NOs:37, 38, and 39, respectively).

- 18 -

Fig. 14 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep222 polypeptide and gene from a *Streptococcus pneumonia* (SEQ ID NOs:40, 41, and 42, respectively).

Fig. 15 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep2283 polypeptide and gene from a *Streptococcus pneumonia* (SEQ ID NOs:43, 44, and 45, respectively).

Fig. 16 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep273 polypeptide and gene from a *Streptococcus pneumonia* strain (SEQ ID NOs:46, 47, and 48, respectively).

Fig. 17 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep286 polypeptide and gene from a *Streptococcus pneumonia* strain (SEQ ID NOs:49, 50, and 51, respectively).

Fig. 18 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep311 polypeptide and gene from a *Streptococcus pneumonia* (SEQ ID NOs:52, 53, and 54, respectively).

Fig. 19 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep3262 polypeptide and gene from a *Streptococcus pneumonia* (SEQ ID NOs:55, 56, and 57, respectively).

Fig. 20 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep3387 polypeptide and gene from a *Streptococcus pneumonia* (SEQ ID NOs:58, 59, and 60, respectively).



- 19 -

Fig. 21 are a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep47 polypeptide and gene from a *Streptococcus pneumonia* strain (SEQ ID NOs:61, 62, and 63, respectively).

Fig. 22 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep61 polypeptide and gene from a *Streptococcus pneumonia* strain (SEQ ID NOs:64, 65, and 66, respectively).

Fig. 23 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the gep76 polypeptide and gene from a *Streptococcus pneumonia* strain (SEQ ID NOs:67, 68, and 69, respectively).

Fig. 24 is a representation of the amino acid and coding strand and non-coding strand nucleic acid sequences of the B-yneS polypeptide and gene from a *Bacillus subtilis* strain (SEQ ID NOs:70, 71, and 72, respectively).

Fig. 25 is a schematic representation of the PCR strategy used to produce DNA molecules used for targeted deletions of essential genes in *Streptococcus pneumoniae*.

Fig. 26 is a schematic representation of the strategy used to produce targeted deletions of essential genes in *Streptococcus pneumoniae*.

### Detailed Description of the Invention

#### Identifying *Streptococcus* Genes in Essential Operons

As shown by the experiments described below, each of the GEP genes is located within an operon that is essential for survival of *Streptococcus pneumonia*. *Streptococcus pneumonia* is available from the ATCC. To identify genes located within essential operons, mutants of *Streptococcus pneumonia* were produced. In

- 20 -

general, mutagenesis of *Streptococcus pneumonia* can be accomplished using any of various art-known methods.

In general, and for the examples set forth below, genes located within essential *Streptococcus pneumonia* operons can be identified using genes from a  
5 *Streptococcus pneumonia* RX1 genomic library, which was produced using standard methods (see Kim et al., Nucl. Acids. Res. 20: 1083-1085 (1992) and Ausubel et al. (eds.), 1995, Current Protocols in Molecular Biology, (John Wiley & Sons, NY)). Genes in this *Streptococcus* library were disrupted using a shuttle mutagenesis approach with the transposon TnPho-A. Each disrupted gene then was  
10 tested to determine whether it was located within an operon that is essential for survival of *Streptococcus pneumonia*. In this method, 2 ml of LB broth supplemented with chloramphenicol (10 µg/ml), MgSO<sub>4</sub> (10 mM) and maltose (0.2%) were inoculated with 50 µl of the *Streptococcus pneumonia* RX-1 plasmid library. The culture was grown at 37°C while shaking until the OD<sub>650</sub> of the  
15 culture reached 0.8 (approximately 2 hours). A 1 ml aliquot of TnPho-A-containing phage (10<sup>9</sup> pfu/ml) was added to 1 ml of the *Streptococcus* culture, producing a ratio of approximately 10 phage to 1 cell. The phage and cells were incubated at 37°C for 30 minutes. A 4 ml aliquot of LB broth, warmed to 37°C, then was added to the phage/cell mixture, and the mixture was incubated at 37°C,  
20 while shaking, for 1 hour. The cells then were pelleted by centrifuging them at 3500 rpm in a Beckman tabletop centrifuge for 5 minutes.

The pelleted cells then were resuspended in 800 µl of LB broth, and a 200 µl aliquot of cells was plated onto each of four petri plates containing LB agar supplemented with chloramphenicol (10 µg/ml), kanamycin (50 µg/ml), and  
25 erythromycin (300 µg/ml). The plates then were incubated overnight at 37°C, and the number of colonies appearing on the plates was counted. Approximately 18,000 colonies then were pooled and used to inoculate 50 ml of LB broth, which was incubated overnight at 37°C. Plasmid DNA from the culture then was extracted using a Qiagen MIDI Prep Kit; other art-known extraction methods can  
30 be substituted.

- 21 -

The concentration of the extracted DNA was measured, and 100 ng of the DNA was transformed, by electroporation, into *E. coli* DH10B cells (Gibco BRL). A 1 ml aliquot of SOC broth then was added the transformed cells, and the cells were incubated at 37°C for 1 hour before being pelleted by centrifugation at 3500  
5 RPM for 5 minutes. The cells then were resuspended in 200 µl of LB broth, and aliquots of 2, 20, and 50 µl were plated onto petri plates containing LB agar and antibiotics as described above. After incubating the plates overnight at 37°C, 93 colonies were picked and used, individually, to inoculate 1.25 ml of Terrific broth supplemented with chloramphenicol (10µg/ml), kanamycin (50µg/ml), and  
10 erythromycin (300µg/ml). The cultures were incubated at 37°C for approximately 20 hours, while shaking. The DNA from each culture then was extracted, using a conventional alkaline lysis miniprep method.

The extracted DNA samples then were used, individually, to transform *Streptococcus pneumonia* cells in a 96-well microtitre format. The transposon  
15 promotes insertion of the mutagenized gene into the bacterial chromosome. Non-transforming clones indicate that the mutation was within an operon containing an essential gene.

The non-transforming clones then were grown in 50 ml of Terrific broth supplemented with chloramphenicol (10 µg/ml), kanamycin (50 µg/ml), and  
20 erythromycin (300 µg/ml). DNA from these clones was extracted and retransformed into *Streptococcus pneumonia* and plated on petri dishes to confirm that they were non-transforming. The genes located within essential operons then were sequenced, using primers that hybridize to sequences of the transposon. The sequences of the primers were: 5'GCAGCCCGGTTTTCCAGAACAGG3' (SEQ ID  
25 NO: 73) and 5'GATTTAGCCCAGTCGGCCGCACG3' (SEQ ID NO: 74).

In an alternative method, which also was used, the transposon Tn 10 was used to disrupt genes in a *Streptococcus pneumonia* fosmid library, which was produced using standard methods. A 50 ml aliquot of TBMM broth supplemented with chloramphenicol (10µg/ml), MgSO<sub>4</sub> (10 mM), and maltose (0.2%) were  
30 inoculated with a single fosmid colony from the fosmid library, and the cultures

- 22 -

were grown overnight at 37°C. The cells then were pelleted and resuspended in 5 ml of LB broth supplemented with chloramphenicol (10 µg/ml), MgSO<sub>4</sub> (10 mM), and maltose (0.2%). A 100 µl aliquot of the cells then was mixed with 100 µl of Tn10 phage lysate (10<sup>10</sup> pfu/ml), and the mixture was incubated at room

5 temperature for 15 minutes and then incubated at 37°C for 15 minutes.

A 5 ml aliquot of LB broth supplemented with IPTG (1 mM) and sodium citrate (50 mM) and warmed to 37°C then was added to the cell/phage mixture.

After incubating the cell/phage mixture at 37°C, while shaking, the cells were pelleted and resuspended in 800 µl of LB broth. The cells then were plated onto 4  
10 plates of LB agar supplemented with chloramphenicol (10 µg/ml) and erythromycin (300 µg/ml). After incubating the cells overnight at 37°C, at least 10,000 of the resulting colonies were used to inoculate 50 ml of LB broth. DNA then was extracted and quantified using standard methods, and 100 ng of DNA were used to transform *E. coli* DH10B cells (Gibco BRL) via electroporation. After adding 1 ml  
15 of SOC broth to the cells, the cells were incubated at 37°C for 1 hour. The cells then were pelleted and suspended in 200 µl LB broth, and aliquots of 2, 20, and 50 µl were plated onto LB agar supplemented with chloramphenicol (10 µg/ml), kanamycin (50 µg/ml), and erythromycin (300 µg/ml). The plates then were incubated overnight at 37°C, and 93 colonies were picked and used to inoculate  
20 1.25 ml of Terrific broth supplemented with chloramphenicol (10 µg/ml), kanamycin (50 µg/ml) and erythromycin (300 µg/ml). These cultures were incubated for approximately 20 hours, while shaking, and the DNA was isolated using a standard miniprep method. The extracted DNA then was used to transform *Streptococcus pneumonia*, and the genes located within essential operons were  
25 sequenced as described above. The sequences of the primers used for sequencing were: 5'CCGCCATTCTTTGCTGTTTCG3' (SEQ ID NO: 75) and 5'TTACACGTTACTAAAGGGAATG3' (SEQ ID NO: 76).

- 23 -

Identification of the gep1493, gep1507, gep1546, gep273, gep286, and gep76 Genes as Essential Genes

As shown by the experiments described below, the gep1493, gep1507, gep1546, gep273, gep286, and gep76 genes each have been shown to be essential  
5 for survival of *Streptococcus pneumoniae*. Each of the gep1493, gep1507, gep1546, gep273, gep286, and gep76 genes has been identified as essential by creating a targeted deletion of each gene, separately, in *Streptococcus pneumoniae*.

Each of the gep1493, gep1507, gep1546, gep273, gep286, and gep76 genes was, separately, replaced with a nucleic acid sequence conferring resistance to the  
10 antibiotic erythromycin (an "erm" gene). Other genetic markers can be used in lieu of this particular antibiotic resistance marker. Polymerase chain reaction (PCR) amplification was used to make a targeted deletion in the *Streptococcus* genomic DNA, as shown in Fig. 25. Several PCR reactions were used to produce the DNA molecules needed to carry out target deletion of the genes of interest. First, using  
15 primers 5 and 6, an erm gene was amplified from pIL252 from *B. subtilis* (available from the *Bacillus* Genetic Stock Center, Columbus, OH). Primer 5 consists of 21 nucleotides that are identical to the promoter region of the erm gene and complementary to Sequence A. Primer 5 has the sequence 5'GTG TTC GTG CTG ACT TGC ACC3' (SEQ ID NO: 77). Primer 6 consists of 21 nucleotides  
20 that are complementary to the 3' end of the erm gene. Primer 6 has the sequence 5'GAA TTA TTT CCT CCC GTT AAA3' (SEQ ID NO: 78). PCR amplification of the erm gene was carried out under the following conditions: 30 cycles of 94°C for 1 minute, 55°C for 1 minute, and 72°C for 1.5 minutes, followed by one cycle of 72°C for 10 minutes.

25 In the second and third PCR reactions, sequences flanking the gene of interest were amplified and produced as hybrid DNA molecules that also contained a portion of the erm gene. The second reaction produced a double-stranded DNA molecule (termed "Left Flanking Molecule") that includes sequences upstream of the 5' end of the gene of interest and the first 21 nucleotides of the erm gene. As  
30 shown in Fig. 25, this reaction utilized primer 1, which is 21 nucleotides in length

- 24 -

and identical to a sequence that is located approximately 500 bp upstream of the translation start site of the gene of interest. Primers 1 and 2 are gene-specific and include the sequences 5'CTC CGT GAA GTC CAC CTG AT3' (SEQ ID NO:79) and 5'GGT GCA AGT CAG CAC GAA CAC GCG ACA TAG GTT CCA GTT  
5 AGG3' (SEQ ID NO:80), respectively, for *gep1493*. Primer 2 is 42 nucleotides in length, with 21 of the nucleotides at the 3' end of the primer being complementary to the 5' end of the sense strand of the gene of interest. The 21 nucleotides at the 5' end of the primer were identical to Sequence A and are therefore complementary to the 5' end of the *erm* gene. Thus, PCR amplification using primers 1 and 2  
10 produced the left flanking DNA molecule, which is a hybrid DNA molecule containing a sequence located upstream of the gene of interest and 21 base pairs of the *erm* gene, as shown in Fig. 25.

The third PCR reaction was similar to the second reaction, but produced the right flanking DNA molecule, shown in Fig. 25. The right flanking DNA molecule  
15 contains 21 base pairs of the 3' end of the *erm* gene, a 21 base pair portion of the 3' end of the gene of interest, and sequences downstream of the gene of interest. This right flanking DNA molecule was produced with gene-specific primers 3 and 4. For *gep 1493*, primers 3 and 4 included the sequences 5'TTT AAC GGG AGG AAA TAA TTC CCA TAT CGT GGC TCC TGA AT 3' (SEQ ID NO:81) and  
20 5'TAA AGC CCT CAT GTC GAA CC3' (SEQ ID NO:82), respectively. Primer 3 is 42 nucleotides; the 21 nucleotides at the 5' end of Primer 3 are identical to Sequence B and therefore are identical to the 3' end of the *erm* gene. The 21 nucleotides at the 3' end of Primer 3 are identical to the 3' end of the gene of interest. Primer 4 is 21 nucleotides in length and is complementary to a sequence  
25 located approximately 500 bp downstream of the gene of interest. As discussed above, primers 1-4 are gene-specific, and the sequences disclosed above were used for *gep1493*. Gene-specific primers were used to identify the other essential genes described herein, as shown in Table 2.

TABLE 2: Primers Used in Identifying Essential Genes

Gene	Primer 1	Primer 2	Primer 3	Primer 4
gep1493	5'CTCCGTGAA GTCCACCTGA T3' (SEQ ID NO:79)	5'GGTGCAAGT CAGCACGAAC ACTGCTCGCG TAGATTGATT TG3' (SEQ ID NO:80)	5'TTTAACGGG AGGAAATAAT TCGGGGATTG AACCTAACCC AT3' (SEQ ID NO:81)	5'TTGGCAAG AAGGCAGAG AAT3' (SEQ ID NO:82)
gep1507	5'GCATGAGAA ACCCAGTCTC C3' (SEQ ID NO:83)	5'GGTGCAAGT CAGCACGAAC ACGCGACATA GGTTCCAGTT AGG3' (SEQ ID NO:84)	5'TTTAACGGG AGGAAATAAT TCCCATATCG TGGCTCCTGA AT3' (SEQ ID NO:85)	5'TAAAGCCC TCATGTCGAA CC3' (SEQ ID NO:86)
gep1546	5'CAGTGACGA TACAGATGAA GAA3' (SEQ ID NO:87)	5'GGTGCAAGT CAGCACGAAC ACGATGCTGG CTTCGTTGAG TG3' (SEQ ID NO:88)	5'TTTAACGGG AGGAAATAAT TCGTCGCGAC TCCTAGCCAT AC3' (SEQ ID NO:89)	5'CCAGCAAA GGAAAACCG ATA3' (SEQ ID NO:90)
gep273	5'GGTCAGTGA CAGCAGCAGA T3' (SEQ ID NO:91)	5'GGTGCAAGT CAGCACGAAC ACGGCCTTGG AAAAAAGACC AT3' (SEQ ID NO:92)	5'TTTAACGGG AGGAAATAAT TCCCGCTTAA ATTCTGCCAA TC3' (SEQ ID NO:93)	5'CCCATAAC CGTATCACCT GG3' (SEQ ID NO:94)
gep286	5'CGGAACGGC TATGAAAAAA A3' (SEQ ID NO:95)	5'GGTGCAAGT CAGCACGAAC ACACGACGAA AGGCAACCAT AC3' (SEQ ID NO:96)	5'TTTAACGGG AGGAAATAAT TCTGGTATGG GGGTTGATGA AG3' (SEQ ID NO:97)	5'TCGCCCTAC TTTTCGTATG C3' (SEQ ID NO:98)
gep76	5'AGCGATATT AGTGCGGGAG A3' (SEQ ID NO:99)	5'GGTGCAAGT CAGCACGAAC ACCAGCAATT TTGTCATCAG TCG3' (SEQ ID NO:100)	5'TTTAACGGG AGGAAATAAT TCCTGGGGTA ATGGAGCACA GT3' (SEQ ID NO:101)	5'GGGATTGT CACGGTAAA ACC3' (SEQ ID NO:102)

- 26 -

PCR amplification of the left and right flanking DNA molecules was carried out, separately, in 50  $\mu$ l reaction mixtures containing: 1  $\mu$ l *Streptococcus pneumoniae* (RX1) DNA (0.25  $\mu$ g), 2.5  $\mu$ l Primer 1 or Primer 4 (10 pmol/ $\mu$ l), 2.5  $\mu$ l Primer 2 or Primer 3 (20 pmol/ $\mu$ l), 1.2  $\mu$ l a mixture dNTPS (10 mM each),  
5 37  $\mu$ l H<sub>2</sub>O, 0.7  $\mu$ l Taq polymerase (5 U/ $\mu$ l), and 5  $\mu$ l 10x Taq polymerase buffer (10 mM Tris, 50 mM KCl, 2.5 mM MgCl<sub>2</sub>). The left and right flanking DNA molecules were amplified using the following PCR cycling program: 95°C for 2 minutes; 72°C for 1 minute; 94°C for 30 seconds; 49°C for 30 seconds; 72°C for 1 minute; repeating the 94°C, 49°C, and 72°C incubations 30 times; 72°C for 10  
10 minutes and then stopping the reactions. A 15  $\mu$ l aliquot of each reaction mixture then was electrophoresed through a 1.2% low melting point agarose gel in TAE buffer and then stained with ethidium bromide. Fragments containing the amplified left and right flanking DNA molecules were excised from the gel and purified using the QIAQUICK™ gel extraction kit (Qiagen, Inc.) Other art-known methods  
15 for amplifying and isolating DNA can be substituted. The flanking left and right DNA fragments were eluted into 30  $\mu$ l TE buffer at pH 8.0.

The amplified *erm* gene and left and right flanking DNA molecules were then fused together to produce the fusion product, as shown in Fig. 25. The fusion PCR reaction was carried out in a volume of 50  $\mu$ l containing: 2  $\mu$ l of each of the  
20 left and right flanking DNA molecules and the *erm* gene PCR product; 5  $\mu$ l of 10x buffer; 2.5  $\mu$ l of Primer 1 (10 pmol/ $\mu$ l); 2.5  $\mu$ l of Primer 4 (10 pmol/ $\mu$ l), 1.2  $\mu$ l dNTP mix (10 mM each) 32  $\mu$ l H<sub>2</sub>O, and 0.7  $\mu$ l Taq polymerase. The PCR reaction was carried out using the following cycling program: 95°C for 2 minutes; 72°C for 1 minute; 94°C for 30 seconds, 48°C for 30 seconds; 72°C for 3 minutes;  
25 repeat the 94°C, 48°C and 72°C incubations 25 times; 72°C for 10 minutes. After the reaction was stopped, a 12  $\mu$ l aliquot of the reaction mixture was electrophoresed through an agarose gel to confirm the presence of a final product of approximately 2 kb.

A 5  $\mu$ l aliquot of the fusion product was used to transform *S. pneumoniae*  
30 grown on a medium containing erythromycin in accordance with standard



- 27 -

techniques. As shown in Fig. 26, the fusion product and the *S. pneumoniae* genome undergo a homologous recombination event so that the *erm* gene replaces the chromosomal copy of the gene of interest, thereby creating a gene knockout. Disruption of an essential gene results in no growth on a medium containing  
5 erythromycin. Using this gene knockout method, the gep1493, gep1507, gep1546, gep273, gep286, and gep76 genes were each identified as being essential for survival.

- 28 -

#### Identification of Homologs and Orthologs of GEP Polypeptides

Having shown that the various GEP genes are essential or located within operons that are essential for survival of *Streptococcus*, it can be expected that homologs and orthologs of the polypeptides encoded by these genes, when present

5 in other organisms, for example *B. subtilis*, are essential or located within operons that are essential for survival of that organism as well, and therefore are useful targets for identifying antibacterial agents. Using the sequences of the GEP polypeptides identified in *Streptococcus*, homologs and orthologs of these polypeptides can be identified in other organisms. For example, the coding

10 sequences of the GEP nucleic acids can be used to search the GenBank database of nucleotide sequences to identify homologs or orthologs that are expressed from essential operons in other organisms. Sequence comparisons can be performed using the Basic Local Alignment Search Tool (BLAST) (Altschul et al., *J. Mol. Biol.*, 215:403-410 1990). The percent sequence identity shared by the GEP

15 polypeptides and their homologs or orthologs can be determined using the GAP program from the Genetics Computer Group (GCG) Wisconsin Sequence Analysis Package (Wisconsin Package Version 9.0, GCG; Madison, WI). The following parameters are suitable: gap creation penalty, 12 (protein) 50 (DNA); gap extension penalty, 4 (protein) 3 (DNA). Typically, the GEP polypeptides and their

20 homologs share at least 25% (e.g., at least 40%) sequence identity. Typically, the DNA sequences encoding GEP polypeptides and their homologs share at least 35% (e.g., at least 45%) sequence identity. To confirm that the homologs or orthologs of the GEP polypeptides are expressed from operons that are essential for survival of bacteria, the operon encoding each of the homologs or orthologs can be,

25 separately, deleted from the genome of the host organism.

#### Identification of Essential Operons in Additional *Streptococcus* Strains

Now that the various GEP genes have been identified as being located within operons that are essential for survival, these genes, or fragments thereof, can be used to detect homologous or orthologous genes in other organisms. In

- 29 -

particular, these genes can be used to analyze various pathogenic and non-pathogenic strains of bacteria. Fragments of a nucleic acid (DNA or RNA) encoding a GEP polypeptide or homolog or ortholog (or sequences complementary thereto) can be used as probes in conventional nucleic acid hybridization assays of pathogenic bacteria. For example, nucleic acid probes (which typically are 8-30, or usually 15-20, nucleotides in length) can be used to detect GEP genes or homologs or orthologs thereof in art-known molecular biology methods, such as Southern blotting, Northern blotting, dot or slot blotting, PCR amplification methods, colony hybridization methods, and the like. Typically, an oligonucleotide probe based on the nucleic acid sequences described herein, or fragments thereof, is labeled and used to screen a genomic library constructed from mRNA obtained from a *Streptococcus* or bacterial strain of interest. A suitable method of labeling involves using polynucleotide kinase to add  $^{32}\text{P}$ -labeled ATP to the oligonucleotide used as the probe. This method is well known in the art, as are several other suitable methods (e.g., biotinylation and enzyme labeling).

Hybridization of the oligonucleotide probe to the library, or other nucleic acid sample, typically is performed under stringent to highly stringent conditions. Nucleic acid duplex or hybrid stability is expressed as the melting temperature or  $T_m$ , which is the temperature at which a probe dissociates from a target DNA. This melting temperature is used to define the required stringency conditions. If sequences are to be identified that are related and substantially identical to the probe, rather than identical, then it is useful to first establish the lowest temperature at which only homologous hybridization occurs with a particular concentration of salt (e.g., SSC or SSPE). Then, assuming that 1% mismatching results in a  $1^\circ\text{C}$  decrease in the  $T_m$ , the temperature of the final wash in the hybridization reaction is reduced accordingly (for example, if sequences having  $\geq 95\%$  identity with the probe are sought, the final wash temperature is decreased by  $5^\circ\text{C}$ ). In practice, the change in  $T_m$  can be between  $0.5^\circ$  and  $1.5^\circ\text{C}$  per 1% mismatch.

As used herein, highly stringent conditions refer to hybridization at  $68^\circ\text{C}$  in 5x SSC/5x Denhardt's solution/1.0% SDS, and washing in 0.2x SSC/0.1% SDS at

- 30 -

42°C. Stringent conditions refer to washing in 3x SSC at 42°C. The parameters of salt concentration and temperature can be varied to achieve the optimal level of identity between the probe and the target nucleic acid. Additional guidance regarding such conditions is readily available in the art, for example, by Sambrook  
5 et al., 1989, *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, N.Y.; and Ausubel et al. (eds.), 1995, *Current Protocols in Molecular Biology*, (John Wiley & Sons, N.Y.) at Unit 2.10.

In one approach, libraries constructed from pathogenic or non-pathogenic *Streptococcus* or bacterial strains can be screened. For example, such strains can  
10 be screened for expression of GEP genes by Northern blot analysis. Upon detection of transcripts of the GEP genes or homologs or orthologs thereof, libraries can be constructed from RNA isolated from the appropriate strain, utilizing standard techniques well known to those of skill in the art. Alternatively, a total genomic DNA library can be screened using an GEP gene probe (or a probe  
15 directed to a homolog or ortholog thereof).

New gene sequences can be isolated, for example, by performing PCR using two degenerate oligonucleotide primer pools designed on the basis of nucleotide sequences within the GEP genes, or their homologs or orthologs, as depicted herein. The template for the reaction can be DNA obtained from strains known or  
20 suspected to express a GEP allele or an allele of a homolog or ortholog thereof. The PCR product can be subcloned and sequenced to ensure that the amplified sequences represent the sequences of a new GEP nucleic acid sequence, or a sequence of a homolog or ortholog thereof.

Synthesis of the various GEP polypeptides or their homologs or orthologs  
25 (or an antigenic fragment thereof) for use as antigens, or for other purposes, can readily be accomplished using any of the various art-known techniques. For example, a polypeptide or homolog or ortholog thereof, or an antigenic fragment(s), can be synthesized chemically *in vitro*, or enzymatically (e.g., by *in vitro* transcription and translation). Alternatively, the gene can be expressed in, and the  
30 polypeptide purified from, a cell (e.g., a cultured cell) by using any of the

- 31 -

numerous, available gene expression systems. For example, the polypeptide antigen can be produced in a prokaryotic host (e.g., *E. coli* or *B. subtilis*) or in eukaryotic cells, such as yeast cells or insect cells (e.g., by using a baculovirus-based expression vector).

- 5 Proteins and polypeptides can also be produced in plant cells, if desired. For plant cells viral expression vectors (e.g., cauliflower mosaic virus and tobacco mosaic virus) and plasmid expression vectors (e.g., Ti plasmid) are suitable. Such cells are available from a wide range of sources (e.g., the American Type Culture Collection, Rockland, MD; also, *see*, e.g., Ausubel et al., *Current Protocols in*
- 10 *Molecular Biology*, John Wiley & Sons, New York, 1994). The optimal methods of transformation or transfection and the choice of expression vehicle will depend on the host system selected. Transformation and transfection methods are described, e.g., in Ausubel et al., *supra*; expression vehicles may be chosen from those provided, e.g., in *Cloning Vectors: A Laboratory Manual* (P.H. Pouwels et
- 15 al., 1985, Supp. 1987). The host cells harboring the expression vehicle can be cultured in conventional nutrient media, adapted as needed for activation of a chosen gene, repression of a chosen gene, selection of transformants, or amplification of a chosen gene.

- If desired, GEP polypeptides or their homologs or orthologs can be
- 20 produced as fusion proteins. For example, the expression vector pUR278 (Ruther et al., *EMBO J.*, 2:1791, 1983) can be used to create *lacZ* fusion proteins. The art-known pGEX vectors can be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can be easily purified from lysed cells by adsorption to glutathione-
- 25 agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target gene product can be released from the GST moiety.

- In an exemplary insect cell expression system, a baculovirus such as *Autographa californica* nuclear polyhedrosis virus (AcNPV), which grows in
- 30 *Spodoptera frugiperda* cells, can be used as a vector to express foreign genes. A

- 32 -

coding sequence encoding a GEP polypeptide or homolog or ortholog can be cloned into a non-essential region (for example the polyhedrin gene) of the viral genome and placed under control of a promoter, e.g., the polyhedrin promoter or an exogenous promoter. Successful insertion of a gene encoding a GEP  
5 polypeptide or homolog or ortholog can result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (i.e., virus lacking the proteinaceous coat encoded by the polyhedrin gene). These recombinant viruses are then used to infect insect cells (e.g., *Spodoptera frugiperda* cells) in which the inserted gene is expressed (see, e.g., Smith et al., *J. Virol.*, 46:584, 1983; Smith,  
10 U.S. Patent No. 4,215,051).

In mammalian host cells, a number of viral-based expression systems can be utilized. When an adenovirus is used as an expression vector, the nucleic acid sequence encoding the GEP polypeptide or homolog or ortholog can be ligated to an adenovirus transcription/ translation control complex, e.g., the late promoter and  
15 tripartite leader sequence. This chimeric gene can then be inserted into the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion into a non-essential region of the viral genome (e.g., region E1 or E3) will result in a recombinant virus that is viable and capable of expressing a essential gene product in infected hosts (see, e.g., Logan, Proc. Natl. Acad. Sci. USA, 81:3655, 1984).

20 Specific initiation signals may be required for efficient translation of inserted nucleic acid sequences. These signals include the ATG initiation codon and adjacent sequences. In general, exogenous translational control signals, including, perhaps, the ATG initiation codon, should be provided. Furthermore, the initiation codon must be in phase with the reading frame of the desired coding  
25 sequence to ensure translation of the entire sequence. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, or transcription terminators (Bittner et al., *Methods in Enzymol.*, 153:516, 1987).

- 33 -

The GEP polypeptides and homologs and orthologs can be expressed individually or as fusions with a heterologous polypeptide, such as a signal sequence or other polypeptide having a specific cleavage site at the N-and/or C-terminus of the protein or polypeptide. The heterologous signal sequence selected  
5 should be one that is recognized and processed, i.e., cleaved by a signal peptidase, by the host cell in which the fusion protein is expressed.

A host cell can be chosen that modulates the expression of the inserted sequences, or modifies and processes the gene product in a specific, desired fashion. Such modifications and processing (e.g., cleavage) of protein products  
10 may facilitate optimal functioning of the protein. Various host cells have characteristic and specific mechanisms for post-translational processing and modification of proteins and gene products. Appropriate cell lines or host systems familiar to those of skill in the art of molecular biology can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this  
15 end, eukaryotic host cells that possess the cellular machinery for proper processing of the primary transcript, and phosphorylation of the gene product can be used. Such mammalian host cells include, but are not limited to, CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, and choroid plexus cell lines.

If desired, the GEP polypeptide or homolog or ortholog thereof can be  
20 produced by a stably-transfected mammalian cell line. A number of vectors suitable for stable transection of mammalian cells are available to the public, *see*, e.g., Pouwels et al. (*supra*); methods for constructing such cell lines are also publicly known, e.g., in Ausubel et al. (*supra*). In one example, DNA encoding the protein is cloned into an expression vector that includes the dihydrofolate reductase  
25 (DHFR) gene. Integration of the plasmid and, therefore, the GEP polypeptide-encoding gene into the host cell chromosome is selected for by including 0.01-300  $\mu$ M methotrexate in the cell culture medium (as described in Ausubel et al., *supra*). This dominant selection can be accomplished in most cell types.

Recombinant protein expression can be increased by DHFR-mediated  
30 amplification of the transfected gene. Methods for selecting cell lines bearing gene

- 34 -

amplifications are described in Ausubel et al. (supra); such methods generally involve extended culture in medium containing gradually increasing levels of methotrexate. DHFR-containing expression vectors commonly used for this purpose include pCVSEII-DHFR and pAdD26SV(A) (described in Ausubel et al.,

5 supra).

A number of other selection systems can be used, including but not limited to, herpes simplex virus thymidine kinase genes, hypoxanthine-guanine phosphoribosyl-transferase genes, and adenine phosphoribosyltransferase genes, which can be employed in *tk*, *hgprt*, or *aprt* cells, respectively. In addition, *gpt*,  
10 which confers resistance to mycophenolic acid (Mulligan et al., *Proc. Natl. Acad. Sci. USA*, 78:2072, 1981); *neo*, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin et al., *J. Mol. Biol.*, 150:1, 1981); and *hygro*, which confers resistance to hygromycin (Santerre et al., *Gene*, 30:147, 1981), can be used.

Alternatively, any fusion protein can be readily purified by utilizing an  
15 antibody or other molecule that specifically binds to the fusion protein being expressed. For example, a system described in Janknecht et al., *Proc. Natl. Acad. Sci. USA*, 88:8972 (1981), allows for the ready purification of non-denatured fusion proteins expressed in human cell lines. In this system, the gene of interest is subcloned into a vaccinia recombination plasmid such that the gene's open reading  
20 frame is translationally fused to an amino-terminal tag consisting of six histidine residues. Extracts from cells infected with recombinant vaccinia virus are loaded onto Ni<sup>2+</sup> nitriloacetic acid-agarose columns, and histidine-tagged proteins are selectively eluted with imidazole-containing buffers.

Alternatively, a GEP polypeptide or homolog or ortholog, or a portion  
25 thereof, can be fused to an immunoglobulin Fc domain. Such a fusion protein can be readily purified using a protein A column, for example. Moreover, such fusion proteins permit the production of a chimeric form of a GEP polypeptide or homolog or ortholog having increased stability *in vivo*.

Once the recombinant GEP polypeptide (or homolog or ortholog) is  
30 expressed, it can be isolated (i.e., purified). Secreted forms of the polypeptides can



- 35 -

be isolated from cell culture media, while non-secreted forms must be isolated from the host cells. Polypeptides can be isolated by affinity chromatography. For example, an anti-gep103 antibody (e.g., produced as described herein) can be attached to a column and used to isolate the protein. Lysis and fractionation of

5 cells harboring the protein prior to affinity chromatography can be performed by standard methods (*see*, e.g., Ausubel et al., supra). Alternatively, a fusion protein can be constructed and used to isolate a GEP polypeptide (e.g., a gep103-maltose binding fusion protein, a gep-103- $\beta$ -galactosidase fusion protein, or a gep103-trpE fusion protein; *see*, e.g., Ausubel et al., supra; New England Biolabs Catalog,

10 Beverly, MA). The recombinant protein can, if desired, be further purified, e.g., by high performance liquid chromatography using standard techniques (*see*, e.g., Fisher, *Laboratory Techniques In Biochemistry And Molecular Biology*, eds., Work and Burdon, Elsevier, 1980).

Given the amino acid sequences described herein, polypeptides useful in

15 practicing the invention, particularly fragments of GEP polypeptides can be produced by standard chemical synthesis (e.g., by the methods described in *Solid Phase Peptide Synthesis*, 2nd ed., The Pierce Chemical Co., Rockford, IL, 1984) and used as antigens, for example.

### Antibodies

20 The GEP polypeptides (or antigenic fragments or analogs of such polypeptides) can be used to raise antibodies useful in the invention, and such polypeptides can be produced by recombinant or peptide synthetic techniques (*see*, e.g., *Solid Phase Peptide Synthesis*, supra; Ausubel et al., supra). Likewise, antibodies can be raised against the GEP homologs and orthologs. In general, the

25 polypeptides can be coupled to a carrier protein, such as KLH, as described in Ausubel et al., supra, mixed with an adjuvant, and injected into a host mammal. Antibodies can be purified, for example, by affinity chromatography methods in which the polypeptide antigen is immobilized on a resin.

- 36 -

In particular, various host animals can be immunized by injection of a polypeptide of interest. Examples of suitable host animals include rabbits, mice, guinea pigs, and rats. Various adjuvants can be used to increase the immunological response, depending on the host species, including but not limited to Freund's  
5 (complete and incomplete adjuvant), adjuvant mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, BCG (bacille Calmette-Guerin) and *Corynebacterium parvum*. Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the  
10 sera of the immunized animals.

Antibodies useful in the invention include monoclonal antibodies, polyclonal antibodies, humanized or chimeric antibodies, single chain antibodies, Fab fragments, F(ab')<sub>2</sub> fragments, and molecules produced using a Fab expression library.

15 Monoclonal antibodies (mAbs), which are homogeneous populations of antibodies to a particular antigen, can be prepared using the GEP polypeptides or homologs or orthologs thereof and standard hybridoma technology (see, e.g., Kohler et al., *Nature*, 256:495, 1975; Kohler et al., *Eur. J. Immunol.*, 6:511, 1976; Kohler et al., *Eur. J. Immunol.*, 6:292, 1976; Hammerling et al., In Monoclonal  
20 Antibodies and T Cell Hybridomas, Elsevier, NY, 1981; Ausubel et al., supra).

In particular, monoclonal antibodies can be obtained by any technique that provides for the production of antibody molecules by continuous cell lines in culture, such as those described in Kohler et al., *Nature*, 256:495, 1975, and U.S. Patent No. 4,376,110; the human B-cell hybridoma technique (Kosbor et al.,  
25 *Immunology Today*, 4:72, 1983; Cole et al., *Proc. Natl. Acad. Sci. USA*, 80:2026, 1983); and the EBV-hybridoma technique (Cole et al., Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96, 1983). Such antibodies can be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD, and any subclass thereof. The hybridomas producing the mAbs of this invention can be cultivated *in*  
30 *vitro* or *in vivo*.

- 37 -

Once produced, polyclonal or monoclonal antibodies are tested for specific recognition of a GEP polypeptide or homolog or ortholog thereof in an immunoassay, such as a Western blot or immunoprecipitation analysis using standard techniques, e.g., as described in Ausubel et al., supra. Antibodies that  
5 specifically bind to the GEP polypeptides, or conservative variants and homologs or orthologs thereof, are useful in the invention. For example, such antibodies can be used in an immunoassay to detect a GEP polypeptide in pathogenic or non-pathogenic strains of bacteria.

Preferably, antibodies of the invention are produced using fragments of the  
10 GEP polypeptides that appear likely to be antigenic, by criteria such as high frequency of charged residues. In one specific example, such fragments are generated by standard techniques of PCR, and are then cloned into the pGEX expression vector (Ausubel et al., supra). Fusion proteins are expressed in *E. coli* and purified using a glutathione agarose affinity matrix as described in Ausubel, et  
15 al., supra.

If desired, several (e.g., two or three) fusions can be generated for each protein, and each fusion can be injected into at least two rabbits. Antisera can be raised by injections in a series, typically including at least three booster injections. Typically, the antisera is checked for its ability to immunoprecipitate a recombinant  
20 GEP polypeptide or homolog or ortholog, or unrelated control proteins, such as glucocorticoid receptor, chloramphenicol acetyltransferase, or luciferase.

Techniques developed for the production of "chimeric antibodies" (Morrison et al., *Proc. Natl. Acad. Sci.*, **81**:6851, 1984; Neuberger et al., *Nature*, **312**:604, 1984; Takeda et al., *Nature*, **314**:452, 1984) can be used to splice the genes from a  
25 mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region.

- 38 -

Alternatively, techniques described for the production of single chain antibodies (U.S. Patent 4,946,778; and U.S. Patents 4,946,778 and 4,704,692) can be adapted to produce single chain antibodies against a GEP polypeptide or homolog or ortholog. Single chain antibodies are formed by linking the heavy and  
5 light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain polypeptide.

Antibody fragments that recognize and bind to specific epitopes can be generated by known techniques. For example, such fragments can include but are not limited to F(ab')<sub>2</sub> fragments, which can be produced by pepsin digestion of the  
10 antibody molecule, and Fab fragments, which can be generated by reducing the disulfide bridges of F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries can be constructed (Huse et al., *Science*, 246:1275, 1989) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

Polyclonal and monoclonal antibodies that specifically bind to GEP  
15 polypeptides or homologs or orthologs can be used, for example, to detect expression of a GEP gene or homolog or ortholog in another strain of bacteria. For example, a GEP polypeptide can be readily detected in conventional immunoassays of bacteria cells or extracts. Examples of suitable assays include, without limitation, Western blotting, ELISAs, radioimmune assays, and the like.

## 20 Assay for Antibacterial Agents

The invention provides a method for identifying an antibacterial agent(s). Although the inventors are not bound by any particular theory as to the biological mechanism involved, the new antibacterial agents are thought to inhibit specifically  
(1) the function of a polypeptide(s) encoded by a nucleic acid located within an  
25 operon containing a GEP gene, or (2) expression of the a gene located within an operon containing a GEP gene, or homologs or orthologs thereof. Screening for antibacterial agents can be rapidly accomplished by identifying those compounds (e.g., polypeptides or small molecules) that specifically bind to a polypeptide encoded by a nucleic acid located within an operon containing a GEP gene. A

- 39 -

homolog or ortholog of a GEP polypeptide can be substituted for the GEP polypeptide in the methods summarized herein. Specific binding of a test compound to a polypeptide can be detected, for example, *in vitro* by reversibly or irreversibly immobilizing the test compound(s) on a substrate, e.g., the surface of a well of a 96-well polystyrene microtitre plate. Methods for immobilizing polypeptides and other small molecules are well known in the art. For example, the microtitre plates can be coated with a polypeptide encoded by a nucleic acid located within an operon containing a GEP gene (e.g., a GEP polypeptide or a combination of GEP polypeptides and/or homologs and/or orthologs) by adding the polypeptide(s) in a solution (typically, at a concentration of 0.05 to 1 mg/ml in a volume of 1-100  $\mu$ l) to each well, and incubating the plates at room temperature to 37°C for 0.1 to 36 hours. Polypeptides that are not bound to the plate can be removed by shaking the excess solution from the plate, and then washing the plate (once or repeatedly) with water or a buffer. Typically, the polypeptide, homolog, or ortholog is contained in water or a buffer. The plate is then washed with a buffer that lacks the bound polypeptide. To block the free protein-binding sites on the plates, the plates are blocked with a protein that is unrelated to the bound polypeptide. For example, 300  $\mu$ l of bovine serum albumin (BSA) at a concentration of 2 mg/ml in Tris-HCl is suitable. Suitable substrates include those substrates that contain a defined cross-linking chemistry (e.g., plastic substrates, such as polystyrene, styrene, or polypropylene substrates from Corning Costar Corp. (Cambridge, MA), for example). If desired, a beaded particle, e.g., beaded agarose or beaded sepharose, can be used as the substrate.

Binding of the test compound to the new polypeptides (or homologs or orthologs thereof) can be detected by any of a variety of art-known methods. For example, an antibody that specifically binds to a GEP polypeptide can be used in an immunoassay. If desired, the antibody can be labeled (e.g., fluorescently or with a radioisotope) and detected directly (*see*, e.g., West and McMahon, *J. Cell Biol.* 74:264, 1977). Alternatively, a second antibody can be used for detection (e.g., a labeled antibody that binds to the Fc portion of an anti-GEP103 antibody).

- 40 -

In an alternative detection method, the GEP polypeptide is labeled, and the label is detected (e.g., by labeling a GEP polypeptide with a radioisotope, fluorophore, chromophore, or the like). In still another method, the GEP polypeptide is produced as a fusion protein with a protein that can be detected optically, e.g.,  
5 green fluorescent protein (which can be detected under UV light). In an alternative method, the polypeptide (e.g., gep103) can be produced as a fusion protein with an enzyme having a detectable enzymatic activity, such as horse radish peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, or glucose oxidase. Genes encoding all of these enzymes have been cloned and are readily available for use by those of skill  
10 in the art. If desired, the fusion protein can include an antigen, and such an antigen can be detected and measured with a polyclonal or monoclonal antibody using conventional methods. Suitable antigens include enzymes (e.g., horse radish peroxidase, alkaline phosphatase, and  $\beta$ -galactosidase), and non-enzymatic polypeptides (e.g., serum proteins, such as BSA and globulins, and milk proteins,  
15 such as caseins).

In various *in vivo* methods for identifying polypeptides that bind to GEP polypeptides, the conventional two-hybrid assays of protein/protein interactions can be used (*see e.g.*, Chien et al., *Proc. Natl. Acad. Sci. USA*, 88:9578, 1991; Fields et al., U.S. Pat. No. 5,283,173; Fields and Song, *Nature*, 340:245, 1989; Le Douarin  
20 et al., *Nucleic Acids Research*, 23:876, 1995; Vidal et al., *Proc. Natl. Acad. Sci. USA*, 93:10315-10320, 1996; and White, *Proc. Natl. Acad. Sci. USA*, 93:10001-10003, 1996). Kits for practicing various two-hybrid methods are commercially available (e.g., from Clontech; Palo Alto, CA).

Generally, the two-hybrid methods involve *in vivo* reconstitution of two  
25 separable domains of a transcription factor. The DNA binding domain (DB) of the transcription factor is required for recognition of a chosen promoter. The activation domain (AD) is required for contacting other components of the host cell's transcriptional machinery. The transcription factor is reconstituted through the use of hybrid proteins. One hybrid is composed of the AD and a first protein

- 41 -

of interest. The second hybrid is composed of the DB and a second protein of interest.

Useful reporter genes are those that are operably linked to a promoter which is specifically recognized by the DB. Typically, the two-hybrid system employs  
5 the yeast *Saccharomyces cerevisiae* and reporter genes, the expression of which can be selected under appropriate conditions. Other eukaryotic cells, including mammalian and insect cells, can be used, if desired. The two-hybrid system provides a convenient method for cloning a gene encoding a polypeptide (i.e., a candidate antibacterial agent) that binds to a second, preselected polypeptide (e.g.,  
10 gep103). Typically, though not necessarily, a DNA library is constructed such that randomly generated sequences are fused to the AD, and the protein of interest (e.g., gep103) is fused to the DB.

In such two-hybrid methods, two fusion proteins are produced. One fusion protein contains the GEP polypeptide (or homolog or ortholog thereof) fused to  
15 either a transactivator domain or DNA binding domain of a transcription factor (e.g., of Gal4). The other fusion protein contains a test polypeptide fused to either the DNA binding domain or a transactivator domain of a transcription factor. Once brought together in a single cell (e.g., a yeast cell or mammalian cell), one of the fusion proteins contains the transactivator domain and the other fusion protein  
20 contains the DNA binding domain. Therefore, binding of the GEP polypeptide to the test polypeptide (i.e., candidate antibacterial agent) reconstitutes the transcription factor. Reconstitution of the transcription factor can be detected by detecting expression of a gene (i.e., a reporter gene) that is operably linked to a DNA sequence that is bound by the DNA binding domain of the transcription  
25 factor.

The methods described above can be used for high throughput screening of numerous test compounds to identify candidate antibacterial (or anti-bacterial) agents. Having identified a test compound as a candidate antibacterial agent, the candidate antibacterial agent can be further tested for inhibition of bacterial growth  
30 *in vitro* or *in vivo* (e.g., using an animal, e.g., rodent, model system) if desired.

- 42 -

Using other, art-known variations of such methods, one can test the ability of a nucleic acid (e.g., DNA or RNA) used as the test compound to bind to a polypeptide encoded by a nucleic acid sequence located within an operon containing a GEP gene or homolog or ortholog thereof.

- 5        *In vitro*, further testing can be accomplished by means known to those in the art such as an enzyme inhibition assay or a whole-cell bacterial growth inhibition assay. For example, an agar dilution assay identifies a substance that inhibits bacterial growth. Microtiter plates are prepared with serial dilutions of the test compound; adding to the preparation a given amount of growth substrate; and  
10 providing a preparation of *Streptococcus* cells. Inhibition of growth is determined, for example, by observing changes in optical densities of the bacterial cultures.

- Inhibition of bacterial growth is demonstrated, for example, by comparing (in the presence and absence of a test compound) the rate of growth or the absolute growth of bacterial cells. Inhibition includes a reduction of one of the above  
15 measurements by at least 20% (e.g., at least 25%, 30%, 40%, 50%, 75%, 80%, or 90%).

- Rodent (e.g., murine) and rabbit animal models of streptococcal infections are known to those of skill in the art, and such animal model systems are accepted for screening antibacterial agents as an indication of their therapeutic efficacy in  
20 human patients. In a typical *in vivo* assay, an animal is infected with a pathogenic *Streptococcus* strain, e.g., by inhalation of *Streptococcus pneumoniae*, and conventional methods and criteria are used to diagnose the mammal as being afflicted with streptococcal pneumonia. The candidate antibacterial agent then is administered to the mammal at a dosage of 1-100 mg/kg of body weight, and the  
25 mammal is monitored for signs of amelioration of disease. Alternatively, the test compound can be administered to the mammal prior to infecting the mammal with *Streptococcus*, and the ability of the treated mammal to resist infection is measured. Of course, the results obtained in the presence of the test compound should be compared with results in control animals, which are not treated with the test



- 43 -

compound. Administration of candidate antibacterial agent to the mammal can be carried out as described below, for example.

#### Pharmaceutical Formulations

- Treatment includes administering a pharmaceutically effective amount of a composition containing an antibacterial agent to a subject in need of such treatment, thereby inhibiting bacterial growth in the subject. Such a composition typically contains from about 0.1 to 90% by weight (such as 1 to 20% or 1 to 10%) of an antibacterial agent of the invention in a pharmaceutically acceptable carrier.
- 10 Solid formulations of the compositions for oral administration may contain suitable carriers or excipients, such as corn starch, gelatin, lactose, acacia, sucrose, microcrystalline cellulose, kaolin, mannitol, dicalcium phosphate, calcium carbonate, sodium chloride, or alginic acid. Disintegrators that can be used include, without limitation, micro-crystalline cellulose, corn starch, sodium starch
- 15 glycolate and alginic acid. Tablet binders that may be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Povidone), hydroxypropyl methylcellulose, sucrose, starch, and ethylcellulose. Lubricants that may be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.
- 20 Liquid formulations of the compositions for oral administration prepared in water or other aqueous vehicles may contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations may also include solutions, emulsions, syrups and elixirs containing, together with the active
- 25 compound(s), wetting agents, sweeteners, and coloring and flavoring agents. Various liquid and powder formulations can be prepared by conventional methods for inhalation into the lungs of the mammal to be treated.
- Injectable formulations of the compositions may contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl

- 44 -

carbonate, isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like). For intravenous injections, water soluble versions of the compounds may be administered by the drip method, whereby a pharmaceutical formulation containing the antibacterial agent and a physiologically acceptable excipient is infused. Physiologically acceptable excipients may include, for example, 5% dextrose, 0.9% saline, Ringer's solution or other suitable excipients. Intramuscular preparations, a sterile formulation of a suitable soluble salt form of the compounds can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. A suitable insoluble form of the compound may be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid, (e.g., ethyl oleate).

A topical semi-solid ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, e.g., 5 to 10% in a carrier such as a pharmaceutical cream base. Various formulations for topical use include drops, tinctures, lotions, creams, solutions, and ointments containing the active ingredient and various supports and vehicles.

The optimal percentage of the antibacterial agent in each pharmaceutical formulation varies according to the formulation itself and the therapeutic effect desired in the specific pathologies and correlated therapeutic regimens. Appropriate dosages of the antibacterial agents can readily be determined by those of ordinary skill in the art of medicine by monitoring the mammal for signs of disease amelioration or inhibition, and increasing or decreasing the dosage and/or frequency of treatment as desired. The optimal amount of the antibacterial compound used for treatment of conditions caused by or contributed to by bacterial infection may depend upon the manner of administration, the age and the body weight of the subject and the condition of the subject to be treated. Generally, the antibacterial compound is administered at a dosage of 1 to 100 mg/kg of body weight, and typically at a dosage of 1 to 10 mg/kg of body weight.

- 45 -

### Example

Using the transposon-based mutagenesis methods described above, the *Streptococcus pneumonia* genome was mutagenized, and 23 genes were identified as being located within operons that are essential for survival of *Streptococcus pneumonia*. These genes are listed in Table 1, above, and their nucleic acid and amino acid sequences are represented by SEQ ID NOs:1-69, as shown in Figs. 1-23.

Now that each of these genes is known to be located within an operon that is essential for survival of *Streptococcus*, the polypeptides encoded by nucleic acids located within those operons can be used to identify antibacterial agents by using the assays described herein. Other art-known assays to detect interactions of test compounds with proteins, or to detect inhibition of bacterial growth also can be used with the nucleic acids located within operons containing the GEP genes, and gene products and homologs or orthologs thereof.

### Other Embodiments

The invention also features fragments, variants, analogs, and derivatives of the GEP polypeptides described above that retain one or more of the biological activities of the GEP polypeptides, e.g., as determined in a complementation assay. Also included within the invention are naturally-occurring and non-naturally-occurring allelic variants. Compared with the naturally-occurring GEP gene, sequences depicted in Figs. 1-23, the nucleic acid sequence encoding allelic variants may have a substitution, deletion, or addition of one or more nucleotides. The preferred allelic variants are functionally equivalent to a GEP polypeptide, e.g., as determined in a complementation assay.

It is to be understood that, while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

- 46 -

What is claimed is:

1. An isolated operon comprising a nucleotide sequence, or an allelic variant or homolog of the nucleotide sequence, encoding:
  - a gep103 polypeptide comprising the amino acid sequence of SEQ ID NO:1,  
5 as depicted in Fig. 1;
  - a gep1119 polypeptide comprising the amino acid sequence of SEQ ID NO:4, as depicted in Fig. 2;
  - a gep1122 polypeptide comprising the amino acid sequence of SEQ ID NO:7, as depicted in Fig. 3;
  - 10 a gep1315 polypeptide comprising the amino acid sequence of SEQ ID NO:10, as depicted in Fig. 4;
  - a gep1493 polypeptide comprising the amino acid sequence of SEQ ID NO:13, as depicted in Fig. 5;
  - a gep1507 polypeptide comprising the amino acid sequence of SEQ ID  
15 NO:16, as depicted in Fig. 6;
  - a gep1511 polypeptide comprising the amino acid sequence of SEQ ID NO:19, as depicted in Fig. 7;
  - a gep1518 polypeptide comprising the amino acid sequence of SEQ ID NO:22, as depicted in Fig. 8;
  - 20 a gep1546 polypeptide comprising the amino acid sequence of SEQ ID NO:25, as depicted in Fig. 9;
  - a gep1551 polypeptide comprising the amino acid sequence of SEQ ID NO:28, as depicted in Fig. 10;
  - a gep1561 polypeptide comprising the amino acid sequence of SEQ ID  
25 NO:31, as depicted in Fig. 11;
  - a gep1580 polypeptide comprising the amino acid sequence of SEQ ID NO:34, as depicted in Fig. 12;
  - a gep1713 polypeptide comprising the amino acid sequence of SEQ ID NO:37 as depicted in Fig. 13;

- 47 -

- a gep222 polypeptide comprising the amino acid sequence of SEQ ID NO:40, as depicted in Fig. 14;
- a gep2283 polypeptide comprising the amino acid sequence of SEQ ID NO:43, as depicted in Fig. 15;
- 5 a gep273 polypeptide comprising the amino acid sequence of SEQ ID NO:46, as depicted in Fig. 16;
- a gep286 polypeptide comprising the amino acid sequence of SEQ ID NO:49, as depicted in Fig. 17;
- a gep311 polypeptide comprising the amino acid sequence of SEQ ID
- 10 NO:52, as depicted in Fig. 18;
- a gep3262 polypeptide comprising the amino acid sequence of SEQ ID NO:55, as depicted in Fig. 19;
- a gep3387 polypeptide comprising the amino acid sequence of SEQ ID NO:58, as depicted in Fig. 20;
- 15 a gep47 polypeptide comprising the amino acid sequence of SEQ ID NO:61, as depicted in Fig. 21;
- a gep61 polypeptide comprising the amino acid sequence of SEQ ID NO:64, as depicted in Fig. 22; or
- a gep76 polypeptide comprising the amino acid sequence of SEQ ID NO:67,
- 20 as depicted in Fig. 23.

2. An isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of:

- (1) an operon comprising the sequence of SEQ ID NO:2, as depicted in Fig. 1, or degenerate variants thereof;
- 25 (2) an operon comprising the sequence of SEQ ID NO:2, or degenerate variants thereof, wherein T is replaced by U;
- (3) nucleic acids complementary to (1) and (2);

- 48 -

(4) fragments of (1), (2), and (3) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:1;

(5) an operon comprising the sequence of SEQ ID NO:5, as depicted in Fig. 2, or degenerate variants thereof;

(6) an operon comprising the sequence of SEQ ID NO:5, or degenerate variants thereof, wherein T is replaced by U;

(7) nucleic acids complementary to (5) and (6);

(8) fragments of (5), (6), and (7) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:4;

(9) an operon comprising the sequence of SEQ ID NO:8, as depicted in Fig. 3, or degenerate variants thereof;

(10) an operon comprising the sequence of SEQ ID NO:8, or degenerate variants thereof, wherein T is replaced by U;

(11) nucleic acids complementary to (9) and (10);

(12) fragments of (9), (10), and (11) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:7;

(13) an operon comprising the sequence of SEQ ID NO:11, as depicted in Fig. 4, or degenerate variants thereof;

(14) an operon comprising the sequence of SEQ ID NO:11, or degenerate variants thereof, wherein T is replaced by U;

(15) nucleic acids complementary to (13) and (14); and

(16) fragments of (13), (14), and (15) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:10;

- 49 -

(17) an operon comprising the sequence of SEQ ID NO:14, as depicted in Fig. 5, or degenerate variants thereof;

(18) an operon comprising the sequence of SEQ ID NO:14, or degenerate variants thereof, wherein T is replaced by U;

5 (19) nucleic acids complementary to (17) and (18);

(20) fragments of (17), (18), and (19) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:13;

(21) an operon comprising the sequence of SEQ ID NO:17, as depicted in  
10 Fig. 6, or degenerate variants thereof;

(22) an operon comprising the sequence of SEQ ID NO:17, or degenerate variants thereof, wherein T is replaced by U;

(23) nucleic acids complementary to (21) and (22);

(24) fragments of (21), (22), and (23) that are at least 15 base pairs in  
15 length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:16;

(25) an operon comprising the sequence of SEQ ID NO:20, as depicted in Fig. 7, or degenerate variants thereof;

(26) an operon comprising the sequence of SEQ ID NO:20, or degenerate  
20 variants thereof, wherein T is replaced by U;

(27) nucleic acids complementary to (25) and (26);

(28) fragments of (25), (26), and (27) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:19;

25 (29) an operon comprising the sequence of SEQ ID NO:23, as depicted in Fig. 8, or degenerate variants thereof;

- 50 -

(30) an operon comprising the sequence of SEQ ID NO:23, or degenerate variants thereof, wherein T is replaced by U;

(31) nucleic acids complementary to (29) and (30); and

(32) fragments of (39), (30), and (31) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:22;

(33) an operon comprising the sequence of SEQ ID NO:26, as depicted in Fig. 9, or degenerate variants thereof;

(34) an operon comprising the sequence of SEQ ID NO:26, or degenerate variants thereof, wherein T is replaced by U;

(35) nucleic acids complementary to (33) and (34);

(36) fragments of (33), (34), and (35) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:25;

(37) an operon comprising the sequence of SEQ ID NO:29, as depicted in Fig. 10, or degenerate variants thereof;

(38) an operon comprising the sequence of SEQ ID NO:29, or degenerate variants thereof, wherein T is replaced by U;

(39) nucleic acids complementary to (37) and (38);

(40) fragments of (37), (38), and (39) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:28;

(41) an operon comprising the sequence of SEQ ID NO:32, as depicted in Fig. 11, or degenerate variants thereof;

(42) an operon comprising the sequence of SEQ ID NO:32, or degenerate variants thereof, wherein T is replaced by U;

(43) nucleic acids complementary to (41) and (42);



- 51 -

(44) fragments of (41), (42), and (43) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:31;

(45) an operon comprising the sequence of SEQ ID NO:35, as depicted in Fig. 12, or degenerate variants thereof;

(46) an operon comprising the sequence of SEQ ID NO:35, or degenerate variants thereof, wherein T is replaced by U;

(47) nucleic acids complementary to (45) and (46); and

(48) fragments of (45), (46), and (47) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:34;

(49) an operon comprising the sequence of SEQ ID NO:38, as depicted in Fig. 13, or degenerate variants thereof;

(50) an operon comprising the sequence of SEQ ID NO:38, or degenerate variants thereof, wherein T is replaced by U;

(51) nucleic acids complementary to (49) and (50);

(52) fragments of (49), (50), and (51) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:37;

(53) an operon comprising the sequence of SEQ ID NO:41, as depicted in Fig. 14, or degenerate variants thereof;

(54) an operon comprising the sequence of SEQ ID NO:41, or degenerate variants thereof, wherein T is replaced by U;

(55) nucleic acids complementary to (53) and (54);

(56) fragments of (53), (54), and (55) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:40;

- 52 -

(57) an operon comprising the sequence of SEQ ID NO:44, as depicted in Fig. 15, or degenerate variants thereof;

(58) an operon comprising the sequence of SEQ ID NO:44, or degenerate variants thereof, wherein T is replaced by U;

5 (59) nucleic acids complementary to (57) and (58);

(60) fragments of (57), (58), and (59) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:39;

(61) an operon comprising the sequence of SEQ ID NO:47, as depicted in  
10 Fig. 16, or degenerate variants thereof;

(62) an operon comprising the sequence of SEQ ID NO:47, or degenerate variants thereof, wherein T is replaced by U;

(63) nucleic acids complementary to (61) and (62); and

(64) fragments of (61), (62), and (63) that are at least 15 base pairs in  
15 length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:46;

(65) an operon comprising the sequence of SEQ ID NO:50, as depicted in Fig. 17, or degenerate variants thereof;

(66) an operon comprising the sequence of SEQ ID NO:50, or degenerate  
20 variants thereof, wherein T is replaced by U;

(67) nucleic acids complementary to (65) and (66);

(68) fragments of (65), (66), and (67) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:49;

25 (69) an operon comprising the sequence of SEQ ID NO:53, as depicted in Fig. 18, or degenerate variants thereof;

- 53 -

(70) an operon comprising the sequence of SEQ ID NO:53, or degenerate variants thereof, wherein T is replaced by U;

(71) nucleic acids complementary to (69) and (70);

(72) fragments of (69), (70), and (71) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:52;

(73) an operon comprising the sequence of SEQ ID NO:56, as depicted in Fig. 19, or degenerate variants thereof;

(74) an operon comprising the sequence of SEQ ID NO:56, or degenerate variants thereof, wherein T is replaced by U;

(75) nucleic acids complementary to (73) and (74);

(76) fragments of (73), (74), and (75) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:55;

(77) an operon comprising the sequence of SEQ ID NO:59, as depicted in Fig. 20, or degenerate variants thereof;

(78) an operon comprising the sequence of SEQ ID NO:59, or degenerate variants thereof, wherein T is replaced by U;

(79) nucleic acids complementary to (77) and (78); and

(80) fragments of (77), (78), and (79) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:58;

(81) an operon comprising the sequence of SEQ ID NO:62, as depicted in Fig. 21, or degenerate variants thereof;

(82) an operon comprising the sequence of SEQ ID NO:62, or degenerate variants thereof, wherein T is replaced by U;

(83) nucleic acids complementary to (81) and (82);

- 54 -

(84) fragments of (81), (82), and (83) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:61;

(85) an operon comprising the sequence of SEQ ID NO:65; as depicted in  
5 Fig. 22, or degenerate variants thereof;

(86) an operon comprising the sequence of SEQ ID NO:65, or degenerate variants thereof, wherein T is replaced by U;

(87) nucleic acids complementary to (85) and (86);

(88) fragments of (85), (86), and (87) that are at least 15 base pairs in  
10 length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:66;

(89) an operon comprising the sequence of SEQ ID NO:68, as depicted in Fig. 23, or degenerate variants thereof;

(90) an operon comprising the sequence of SEQ ID NO:68, or degenerate  
15 variants thereof, wherein T is replaced by U;

(91) nucleic acids complementary to (89) and (90); and

(92) fragments of (89), (90), and (91) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:67.

20 3. An isolated operon from *Streptococcus* comprising a nucleotide sequence that is at least 85% identical to a nucleotide sequence selected from the group consisting of

SEQ ID NO:2;

SEQ ID NO:5;

25 SEQ ID NO:8;

SEQ ID NO:11;

SEQ ID NO:14;

- 55 -

5       SEQ ID NO:17;  
       SEQ ID NO:20;  
       SEQ ID NO:23;  
       SEQ ID NO:26;  
       SEQ ID NO:29;  
       SEQ ID NO:32;  
       SEQ ID NO:35;  
       SEQ ID NO:38;  
       SEQ ID NO:41;  
10       SEQ ID NO:44;  
       SEQ ID NO:47;  
       SEQ ID NO:50;  
       SEQ ID NO:53;  
       SEQ ID NO:56;  
15       SEQ ID NO:59;  
       SEQ ID NO:62;  
       SEQ ID NO:65; and  
       SEQ ID NO:68.

4. An isolated nucleic acid molecule that is at least 15 base pairs in length  
20 and hybridizes under stringent conditions to a nucleotide sequence selected from  
the group consisting of

25       SEQ ID NO:2;  
       SEQ ID NO:5;  
       SEQ ID NO:8;  
       SEQ ID NO:11;  
       SEQ ID NO:14;  
       SEQ ID NO:17;  
       SEQ ID NO:20;  
       SEQ ID NO:23;

- 56 -

SEQ ID NO:26;  
SEQ ID NO:29;  
SEQ ID NO:32;  
SEQ ID NO:35;  
5 SEQ ID NO:38;  
SEQ ID NO:41;  
SEQ ID NO:44;  
SEQ ID NO:47;  
SEQ ID NO:50;  
10 SEQ ID NO:53;  
SEQ ID NO:56;  
SEQ ID NO:59;  
SEQ ID NO:62;  
SEQ ID NO:65; and  
15 SEQ ID NO:68.

5. A vector comprising an operon of claim 1.

6. A vector comprising a nucleic acid molecule of claim 2.

7. An expression vector comprising an operon of claim 1 operably linked to a nucleotide sequence regulatory element that controls expression of said operon.

20 8. An expression vector comprising a nucleic acid molecule of claim 2, wherein said nucleic acid molecule is operably linked to a nucleotide sequence regulatory element that controls expression of said nucleic acid.

9. A host cell comprising an exogenously introduced operon of claim 1.

- 57 -

10. A host cell comprising an exogenously introduced nucleic acid molecule of claim 2.

11. A host cell of claim 9, wherein the cell is a yeast or bacterium.

12. A host cell of claim 10, wherein the cell is a yeast or bacterium.

5        13. A genetically engineered host cell comprising an operon of claim 1 operably linked to a heterologous nucleotide sequence regulatory element that controls expression of the operon in the host cell.

14. A host cell of claim 13, wherein the cell is a yeast or bacterium.

10        15. A genetically engineered host cell comprising a nucleic acid molecule of claim 2 operably linked to a nucleotide sequence regulatory element that controls expression of the nucleic acid in the host cell.

16. A host cell of claim 15, wherein the cell is a yeast or bacterium.

15        17. An isolated operon comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of:

the amino acid sequence of SEQ ID NO:1, as depicted in Fig. 1;

the amino acid sequence of SEQ ID NO:4, as depicted in Fig. 2;

the amino acid sequence of SEQ ID NO:7, as depicted in Fig. 3;

the amino acid sequence of SEQ ID NO:10, as depicted in Fig. 4;

20        the amino acid sequence of SEQ ID NO:13, as depicted in Fig. 5;

the amino acid sequence of SEQ ID NO:16, as depicted in Fig. 6;

the amino acid sequence of SEQ ID NO:19, as depicted in Fig. 7;

the amino acid sequence of SEQ ID NO:22, as depicted in Fig. 8;

- 58 -

the amino acid sequence of SEQ ID NO:25, as depicted in Fig. 9;  
the amino acid sequence of SEQ ID NO:28, as depicted in Fig. 10;  
the amino acid sequence of SEQ ID NO:31, as depicted in Fig. 11;  
the amino acid sequence of SEQ ID NO:34, as depicted in Fig. 12;  
5 the amino acid sequence of SEQ ID NO:37, as depicted in Fig. 13;  
the amino acid sequence of SEQ ID NO:40, as depicted in Fig. 14;  
the amino acid sequence of SEQ ID NO:43, as depicted in Fig. 15;  
the amino acid sequence of SEQ ID NO:46, as depicted in Fig. 16;  
the amino acid sequence of SEQ ID NO:49, as depicted in Fig. 17;  
10 the amino acid sequence of SEQ ID NO:52, as depicted in Fig. 18;  
the amino acid sequence of SEQ ID NO:55, as depicted in Fig. 19;  
the amino acid sequence of SEQ ID NO:58, as depicted in Fig. 20;  
the amino acid sequence of SEQ ID NO:61, as depicted in Fig. 21;  
the amino acid sequence of SEQ ID NO:64, as depicted in Fig. 22; and  
15 the amino acid sequence of SEQ ID NO:67, as depicted in Fig. 23.

18. An isolated polypeptide encoded by a nucleic acid located within an operon comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 44, 47, 50, 53, 56, 59, 62, 65, and 68.

20 19. An isolated polypeptide, said polypeptide being encoded by an operon of claim 1.

20. An isolated polypeptide, said polypeptide being encoded by a nucleic acid molecule of claim 2.

21. An isolated polypeptide, said polypeptide being encoded by an  
25 operon of claim 3.



- 59 -

22. A method for identifying an antibacterial agent, the method comprising:

- (a) contacting a test compound with a polypeptide, or a homolog of a polypeptide, encoded by a nucleic acid sequence located within an operon comprising a GEP gene selected from the group consisting of gep103, gep1119, gep1122, gep1315, gep1493, gep1507, gep1511, gep1518, gep1546, gep1551, gep1561, gep1580, gep1713, gep222, gep2283, gep273, gep286, gep311, gep3262, gep3387, gep47, gep61, and gep76; and
- (b) detecting binding of the test compound to the polypeptide, wherein binding indicates that the test compound is an antibacterial agent.

23. The method of claim 22, further comprising:

- (c) determining whether a test compound that binds to the polypeptide inhibits growth of bacteria, relative to growth of bacteria cultured in the absence of a test compound that binds to the polypeptide, wherein inhibition of growth indicates that the test compound is an antibacterial agent.

24. The method of claim 22, wherein the polypeptide is selected from the group consisting of gep103, gep1119, gep1122, gep1315, gep1493, gep1507, gep1511, gep1518, gep1546, gep1551, gep1561, gep1580, gep1713, gep222, gep2283, gep273, gep286, gep311, gep3262, gep3387, gep47, gep61, and gep76.

25. The method of claim 22, wherein the test compound is immobilized on a substrate, and binding of the test compound to the polypeptide is detected as immobilization of the polypeptide on the immobilized test compound.

26. The method of claim 25, wherein immobilization of the polypeptide on the test compound is detected in an immunoassay with an antibody that specifically binds to the polypeptide.

- 60 -

27. The method of claim 22, wherein the test compound is selected from the group consisting of polypeptides and small molecules.

28. The method of claim 22, wherein:

(a) the polypeptide is provided as a fusion protein comprising the  
5 polypeptide fused to (i) a transcription activation domain of a transcription factor or  
(ii) a DNA-binding domain of a transcription factor; and

(b) the test compound is a polypeptide that is provided as a fusion protein  
comprising the test polypeptide fused to (i) a transcription activation domain of a  
transcription factor or (ii) a DNA-binding domain of a transcription factor, to  
10 interact with the fusion protein; and

(c) binding of the test compound to the polypeptide is detected as  
reconstitution of a transcription factor.

29. An antibody that specifically binds to a GEP polypeptide of claim 19.

30. An antibody of claim 29, wherein the antibody is a monoclonal  
15 antibody.

31. A method for identifying an antibacterial agent, the method comprising:

(a) contacting a polypeptide encoded by a nucleic acid located within an  
operon comprising a GEP gene with a test compound;

(b) detecting a decrease in function of the polypeptide contacted with the  
20 test compound; and

(c) determining whether a test compound that decreases function of a  
contacted polypeptide inhibits growth of bacteria, relative to growth of bacteria  
cultured in the absence of a test compound that decreases function of a contacted  
polypeptide, wherein inhibition of growth indicates that the test compound is an  
25 antibacterial agent.

- 61 -

32. The method of claim 31, wherein the polypeptide is selected from the group consisting of gep103, gep1119, gep1122, gep1315, gep1493, gep1507, gep1511, gep1518, gep1546, gep1551, gep1561, gep1580, gep1713, gep222, gep2283, gep273, gep286, gep311, gep3262, gep3387, gep47, gep61, and gep76.

5        33. The method of claim 31, wherein the test compound is selected from the group consisting of polypeptides and small molecules.

34. A method for identifying an antibacterial agent, the method comprising:

(a) contacting a nucleic acid comprising an operon containing a gene encoding a GEP polypeptide with a test compound, wherein the GEP polypeptide is  
10 selected from the group consisting of gep103, gep1119, gep1122, gep1315, gep1493, gep1507, gep1511, gep1518, gep1546, gep1551, gep1561, gep1580, gep1713, gep222, gep2283, gep273, gep286, gep311, gep3262, gep3387, gep47, gep61, and gep76; and

(b) detecting binding of the test compound to the nucleic acid, wherein  
15 binding indicates that the test compound is an antibacterial agent.

35. The method of claim 34, further comprising:

(c) determining whether a test compound that binds to the nucleic acid inhibits growth of bacteria, relative to growth of bacteria cultured in the absence of the test compound that binds to the nucleic acid, wherein inhibition of growth  
20 indicates that the test compound is an antibacterial agent.

36. The method of claim 34, wherein the test compound is selected from the group consisting of polypeptides and small molecules.

37. An isolated nucleic acid or an allelic variant thereof encoding:

a gep1493 polypeptide comprising the amino acid sequence of SEQ ID  
25 NO:13, as depicted in Fig. 5;

- 62 -

a gep1507 polypeptide comprising the amino acid sequence of SEQ ID NO:16, as depicted in Fig. 6;

a gep1546 polypeptide comprising the amino acid sequence of SEQ ID NO:25, as depicted in Fig. 9;

5 a gep273 polypeptide comprising the amino acid sequence of SEQ ID NO:46, as depicted in Fig. 16;

a gep286 polypeptide comprising the amino acid sequence of SEQ ID NO:49, as depicted in Fig. 17; or

a gep76 polypeptide comprising the amino acid sequence of SEQ ID NO:67,  
10 as depicted in Fig. 23.

38. An isolated nucleic acid comprising a sequence selected from the group consisting of:

(1) SEQ ID NO:14, as depicted in Fig. 5, or degenerate variants thereof;

(2) SEQ ID NO:14, or degenerate variants thereof, wherein T is replaced by  
15 U;

(3) nucleic acids complementary to (1) and (2);

(4) fragments of (1), (2), and (3) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:13;

20 (5) SEQ ID NO:17, as depicted in Fig. 6, or degenerate variants thereof;

(6) SEQ ID NO:17, or degenerate variants thereof, wherein T is replaced by U;

(7) nucleic acids complementary to (5) and (6);

(8) fragments of (5), (6), and (7) that are at least 15 base pairs in length and  
25 which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:16;

(9) SEQ ID NO:26, as depicted in Fig. 9, or degenerate variants thereof;

(10) SEQ ID NO:26, or degenerate variants thereof, wherein T is replaced by U;

- 63 -

- (11) nucleic acids complementary to (9) and (10);
- (12) fragments of (9), (10), and (11) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:25;
- 5 (13) SEQ ID NO:47, as depicted in Fig. 16, or degenerate variants thereof;
- (14) SEQ ID NO:47, or degenerate variants thereof, wherein T is replaced by U;
- (15) nucleic acids complementary to (13) and (14);
- (16) fragments of (13), (14), and (15) that are at least 15 base pairs in
- 10 length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:46;
- (17) SEQ ID NO:50, as depicted in Fig. 17, or degenerate variants thereof;
- (18) SEQ ID NO:50, or degenerate variants thereof, wherein T is replaced by U;
- 15 (19) nucleic acids complementary to (i) and (j);
- (20) fragments of (i), (j), and (k) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding the polypeptide of SEQ ID NO:49;
- (21) SEQ ID NO:68, as depicted in Fig. 23, or degenerate variants thereof;
- 20 (22) SEQ ID NO:68, or degenerate variants thereof, wherein T is replaced by U;
- (23) nucleic acids complementary to (21) and (22); and
- (24) fragments of (21), (22), and (23) that are at least 15 base pairs in length and which hybridize under stringent conditions to genomic DNA encoding
- 25 the polypeptide of SEQ ID NO:67.

39. A method for identifying an antibacterial agent, the method comprising:

- (a) contacting a test compound with a polypeptide, or a homolog of a polypeptide, encoded by a nucleic acid sequence located within an operon comprising a B-yneS gene; and

- 64 -

(b) detecting binding of the test compound to the polypeptide, wherein binding indicates that the test compound is an antibacterial agent.

40. The method of claim 39, further comprising:

- (c) determining whether a test compound that binds to the polypeptide
- 5 inhibits growth of bacteria, relative to growth of bacteria cultured in the absence of a test compound that binds to the polypeptide, wherein inhibition of growth indicates that the test compound is an antibacterial agent.

gsp103

Fig. 1

(SEQ ID NO: 2) 1 TGCTGATTTTGGAGAAAGTTTATTAGAGATAAAAGAGTCTAAGGAAAAAATTCATTTCATATTTTCTTCTATAAAATAGATAAAAAATGGTACATA 100  
ACGACTAAAAACCTCTTTCAAATAATCTCTATTTTCTCAGATTCCTTTTAAAGGTAAACTATAAAAAGAAGATATTTATCTATTTTACCATGTTAT

(SEQ ID NO: 3)

101 ATAAATTGAGGTAAATAAGGATGACATTAGATAAATATTTAAAAGTATCGCGAATTATCAAGCGTCGTACAGTCGCAAGGAAGTAGCAGATAAAGGTAGA 200  
TATTTAACTCCATTATTCCTACTCTAATCTATTTATAAATTTTCATAGCGCTTAATAGTTTCGCAGCATGTCAGCGTTTCCTTCATCGTCTATTTCCATCT

(SEQ ID NO: 1) 1 M R L D K Y L K V S R I I K R R T V A K E V A D K G R 27

201 ATCAAGGTTAATGGAATCTTGGCCAAAAGTTCAACCGACTTGAAGTTAATGACCAAGTTGAAATTCGCTTTGGCAATAAGTTGCTGCTTGTAAAAGTAC 300  
TAGTTCCAATTACCTTAGAACCCTTTCAAGTTGCTGAACTTTCAATTACTGTTCAACTTTAAGCGAAACCGTTATTCAACGACGAACATTTTCATG

28 I K V M G I L A K S S T D L K V N D Q V E I R F G M K L L L V K V L 61

101 TAGAGATGAAGATAGTACAAAAAAGAAGATGCAGCAGGAAATGTATGAAATTATCAGTGAAAACACCGGTAGAAGAAAAATGCTAAAAATATTGTACAAT 400  
ATCTCTACTTTCTATCATGTTTTTTCTTCTACGTGTCCTTACATACTTTAATAGTCACTTTGTGCCATCTTCTTTTACAGATTTTATAACATGTTA

62 E M K D S T K K E D A A G M Y E I I S E T R V E E N V . 69

gsep1119

Fig. 2

(SEQ ID NO: 5) 1 GAAATCCGTTTCCAATGTGACTGTAGCCATGAACGGCTTTATGAACGCTCTTCCAGCTCAGACTTACAGGAAATGAAGACGGAACCAACG 100  
(SEQ ID NO: 6) CTTTAGGCAAGGTTACACTGACATCGGTACTTGGAAATACTTTCGGAACGGTTCGGAAGGTTGGAGTCTGAATGCTCTTACTTCTCTCTCTGCTG

101 GGGCAGAAATCACTTGTCAATTCGCCAACTACTTACAACCTTGTATGAAAGGACCTGGAGAACTCATTGGTGACAAATCTTATACACCTTTTATGA 200  
(SEQ ID NO: 4) 1 CCCGTCTTAACTGAAACAGTTAAGACGGTTTGTGAATGTTGAAACTACTTTCCTGGACCTCCTTGAGTAAGCACTGTTTAGAATTAATGCGAAATACT

201 TTGGCAATATTGAGATTCCAATCGTACCGTTTATAGCGCTATGGCTGGCTGACCAACTCAGCCTTTCCTACCATCGCAAAAGAGCTCGGAGCTGCACT 300  
AACCGTTATAACTCTAAGCGTTAGCATGGCAAAATCGCGATACCGACCGCACTGGTTGAGTGGGAAAGCATGGTAGCGTTTTCGAGCGCTCGACCTGA

20 G M I E I P N R T V L A P H A G V T N S A F R T I A K E L G A G L 52

301 CGTGTAAATGGAAATGGTCTCTGACAAAGCAATCCAATCAACCAAGAAACCCCTGCATATGCTTCATATCGATGAGGCGGAAACCCCTCTCTATC 400  
GCAACATTACCTTTACAGAGACTGTTCCCTTAGGTTATGTTGCTTTTTTGGGAGTATACGAAGTATAGCTACTCCGCGTTTGGGACAGAGTAG

53 V V H E M V S D K G I Q Y N N E K T L H M L H I D E G E N P V S I 85

401 CAACTTTTGGTAGCGATGAAGACAGCTAGCACGGCGCAGCAGAAATTCATCCAAGAAACACCAAGACCGATATCGTGGATATCAACATGGGCTGCCCTG 500  
GTTGAAAAACCATCGCTACTTCTGTCGATCGTGGCGCTGCTCTAAGTAGGTTCTTTGTGTTCTGGCTATAGCAGCTATAGTTGTTCCAGGTGAGACAGGAATATAGGG

86 O L F G S D E D S L A R A A E F I O E N T K T D I V D I N M G C P V 119

501 TCAACAAATCGTGAAGAACGAAGCTGGAGCTATGTGGCTCAAGGATCGTGACCAAGATCTACTTATCATCAACAAAGTCCAGTCTGTCTTGTATATCCC 600  
AGTTGTTTATAGCACTTCTTGTCTGACCTCGATACACCGAGTTCTTAGGACTGTTCTAGATGAGATAGTAGTTGTTCCAGGTGAGACAGGAATATAGGG

120 N K I V K N E A G A M W L K D P D K I Y S I I N K V Q S V L D I P 152

601 ACTTACTGTCAAAATGCGTACCGGCTGGCGGACCCATCTTGGCAGTAGAAATGCCCTCGCTGCTGAGGCTGCAGGTGTTTCTGCCCTCGCCATGCAT 700  
TGAATGACAGTTTACGCATGGCGGACCGGCTGGTAGAGACCGTCATCTTTACGGGAGCGGAGCTCCGACGTTCCACAAAGACGGGAGCGGTACGTA

153 L T V K M R T G W A D P S L A V E N A L A A E A A G V S A L A M H 185

701 GCGCGTACCGGTGAACAAATGTATACCTGGCCAGCAGACCTTGAGACCTTTACAAGGTTGCCCAAGCTCTAACCAAGATTCCATTTCATGCCCAACGGTG 800  
CGGCATGGGCACTGTTTACATATGACCGGTGGCTCGGAACCTCTGGGAAATGTTCCAACGGGTTGAGATTGGTTCTAAGGTAAAGTACCGGTGGCCAC

186 G R T R E O M Y T G H A D L E T L Y K V A Q A L T K I P F I A N G D 219

801 ATATCGTACTCTCCAAGAGCCAAAGCAACGATCGAAGAAGTGGTGTGACCGAGTCATGATTGGCGGAGCTGCCATGGGAAATCCTTACCTTTCAA 900  
TATAGGCATGACAGGTTCTTCGGTTCTTGGGTAGCTTCTCAACACGAGCTGCGTCACTAACCAGGCTCGACGGTACCGTTTAGGAAATGGAGAAGTT

220 I R T V Q E A K O R I E E V G A D A V M I G R A A M G N P Y L F N 252

901 CCAATCAACCACTTACTTTGAAACAGGAGAAATCCTACCTGATTGACCTTTGAAGACAAGATGAAGATCGCCTACGAACACTTGAACGATTGATTAA 1000  
CGTTTAGTTGGTAATGAACTTCTCTCTTTAGGATGCACTAACTGGAACTTCTGTTCTACTTCTAGCGGATGCTTGTGAATTTGCTAACTAAATG

253 O : N H Y F E T G E I L P D L T F E D K M K I A Y E H L K R L I N 285

1001 CTCAAAGGAGAAAAAGTCCGAGTTGGTAATTCGGCGGCTCGCTCCTCACTATCTCGGTGGAAACATCTGGCGTGCCTAACTCCGTGGAGCCATTTGCG 1100  
GAGTTTCTCTTTGACGCTCAAGCACTTAAGGCGCGGAGCGAGGAGTGATAGAGGCACCTGTAGACCGGACGGTTTGAAGCACTCGGTAAAGCG

286 L Y G E N V A V R E F R C L A P H Y L R G T S G A A K L R G A I S O 319

1101 AAGCTAGCACCTAGCAGAGATTGAAGCCTCTTCCAATTCGAGAAGCTTAATAGTTTAAAAACCGTAACCTCTTAAAGAGTCTCTTGAATGCCGCCA 1200  
TTCGATCGTGGATCCTCTAACTTCGGGAGACGTTAACTCTTCCAATTAATCAAAATTTGGGCAATGAGAGAAATTTCTCAGAGAACTTACGGCGGT

320 A S T L A E : E A L L O L E K A \* 336



gsep1122

Fig. 3 (Sheet 1 of 2)

(SEQ ID NO: 8) 1 AAGCCAGGAGCTGGAAGTTTCCCTCATATTTTCAATAGTTTATTAGCTACACGTTGAGCACTTCAGAAAAATCAATTTCTTCAAGTTCTCTCTCTA 100  
(SEQ ID NO: 9) TCCGTGCTCGACCTTCAAAAGGAGTATAAAAAGTTATCAATATCGATGTGCAACTCGTTGAAGTCTTTTAGTTTAAGAAAGTTCAAGAGAGAT

101 TAGTAGATTTGAAATCCCTTTTTCAGCTAGTTCTGAGTCAGCACATAAGGACCTTGTCTCTGAAAGTTGATTGGTATTCATGATAGCATAGCGTA 200  
ATCATCTAAAACCTTAGGGAAAACTCGATCAAGACTCAGTCGTGATTCTCTGGGAAACAGAGGACTTCAACTAACCATAACTACTATCGTATTCCAT

201 CTGACCATCATTAATCCACTTATCTTCTTAAGATTAGCAATACTTGAGAAACGATGTTTTATCAATATCGTATTTTTCAGATATTTCTGACTTCT 300  
GACTGGTAGTAATTAGGTGAATAGAGAAATTTCTAATCGTTATTGAATCTTTGCTACAAAAATAGTTATAGCATAAAAAGTCTATAGGAGACTGAAGA

301 TTTTCAGTGGTGCTTTAAAGGATAAGTGGTAGAGGCCAGATTCTTACCATAGAAAAATTCAGCAAGTCTTCAATCTCTTTCAATTCCTCTTCGGTTA 400  
AAAAGTCAGCACGAAATTTCTATTCACTCTCCCGCTAAGAAATGGTATTCTTTTAACTCGTTTCAGAACTTAGAGAAAGTTAAGGAGAGCGAAT

401 TCACCTTATCTCTCGATAACATAAAACGAACTAATGTATCTTCGGTGATAGCATTGTCTGCCATTATCAAGCTCCATCAGATAGAGTCTTTTTTCTT 500  
AGTGGAAATAGAGAGCTATTGTATTTCTTGTAAATAGAGCCATATCGTAAACAGCGGTAAATAGTTGAGGTAGTCTATCTCAGAAAAAAGAA

501 TTCAAGTTTGTGATTTTCATAGCTCTATTATACTCAAAATGTGATAGATAGCGGTATGAATCTGAAAGTGAAACAAAAATACCATTAAAAATCAAG 600  
AAGTTCAAAACACTAAAAGTATCGAGATAATATTGAGTTTACACTATTCTATCCCCATCTTAGACTTTCACCTTTGTTTTATGGTAATTTTAGTTT

(SEQ ID NO: 7) 1 M N L K V K Q K I P L K I X 14

601 CGCATGGGAATTAAAGGTGAGGGAATCGGCTTTTACCAAAAAACATTAGTCTTTGTACCAGGAGCTCTCAAGGCCAAGATATCTATTGTGAGATTACTT 700  
CGGTACCTTAATTCGCACTCCCTTAGCCGAAAAATGGTTTTGTAAATCAGAAACATGGTCTCGAGAGTTTCCGCTCTATAGATAACAGTCTAATGAA

15 R M G : N G E G I G F Y O K T L V F V P G A L K G E D : Y C Q I T S 48

701 CTATTAGACGCAACTTGTGGAAGCAAAATTAAGGTCAACGAAGTCTAAATTTGGAATGTGCCATCTTGTACTATTATAAGTAATGCCAGG 800  
GATAATCTCGTTGAAACAACTTCGTTTAAAGTCTTCAGTCTCTTCAGATTTAAAGCTTAACACGGTAGAAGTGAATAATTAATTAAGTCTCC

49 I R R N F V E A K L L K V N K K S K F R I V P S C T I Y N E C G G 81

801 CTGCCAAATCATGCACCTGCATTATGATAAGCAGTGGAGTCAAGACGCACTTCTCATCAAGCCTGAAAAATTTGCTCTGAGGATATGAAAT 900  
GACGGTTAGTACGTGAGGTAAATTAATCTTCTGACCTCAAGTCTGCTGAAATGAAGTAGTTCGGGACTTTTAAACAGGAGGCTCTATTAATTTTA

82 C Q I M H L H Y D K O L E F K T D L L H O A L K K F A P A G Y E N 114

901 TATGAAATTCGTCAACTATTGGAATGCAGGAACCAAAATATTACAGGCTAAGTTACAATTTAGACTCGAAAAATTTAAAAATCAGGTCAAGGCCGGCT 1000  
ATACTTTAAGCAGGTTGATAACCTTACGCTCTGCTTTTATAATGTCTCGATTCAATGTTAAAGTCTGAGCTTTAAATTTTATGCTCAGTTCCGCCCCG

115 Y E I R P T : G M Q E P K Y Y R A K L O F C T R K F K N Q V K A G L 148

1001 TATATGCACAAAACCTCTCACTATTAGTAGAGTTGAAGACTGCTGCTACAAGATAAGGAAACCAAGTGATTGCTAATCGCTTAGCAGAACTACTTAC 1100  
ATATACGTTTTCAGAGTGATAAATCATCTCACTCTGACGGACCATGTTCTATTCTTTGGGTTCACTAACGATTAGCGAATCGTCTTAATGAATG

149 Y A Q M S H Y L V E L K D C L V O D K E T Q V I A N R L A E L L T 181

1101 TTATCACCAGATTCCAATCAGGATGAGAGAAAACTTCTAGGTGCTGCTACTATTATGGTCCGACGGCGAGAAAGACCGGACAGGTTGAGATTATTAT 1200  
AATAGTGGTCTAAGGTTAGTGCCTACTCTCTTTCAAGATCCACAGGATGATAATACAGGCTGCGGCTCTTCTGGCTGTCCAAGTCTAATAATAA

182 Y H Q I P : T D E R K V L G V R T I M V R R A R K T G Q V Q I I : 214

1201 GTTACAAAACCGCAGCTTAATTAACCTCAATTCGTAAAAGAGTTGGTTAAAGATTTCAGAAAGTTGTGACAGTAGCTGTTAATACAAATACAGCTAAAA 1300  
CAATGTTTGGCGGTCGAATTAATTAAGTTAACCATTCTCAACCAATTTCTAAAGGCTCTCAACAGTGTATGCAAAATTAATGTTATGTCGATTTT

215 V T N R D L N L T O L V K E L V K D F P E V V T V A V N T N T A K T 248

1301 CCAGTGAGATATATGGTGAAGAGACAGAGATTATCTGGGGCAAGAGAGATTCAAGAAAGGTGACTCAATTATGAATTTTCACTATCCCTTCAGGCTTT 1400  
GGTCACTCTATATACCACTTTCTGTCTAATAGACCCCTCTCTCATAACTTCTCCACATGAGTTAATACTTAAAGTGATAGGGAGCTCGAAA

## Fig. 3 (Sheet 2 of 2)

249 S E I Y G E K T E I I M G O E S I O E G V L N Y E F S L S P R A F 281

1401 TTATCAACTAAATCCTGAGCAACAGAACTCCTCTATAGCGAAGCAGTAAAGCGCTGGATGTTGATAAAGAAGACCATTTCATTGACGCTTATTGTGCA 1500  
AATAGTTGATTAGGACTCGTTTGTCTTCAGAGATATCGCTTCGTCATTTTCGGACCTACAACTATTTCTTCTGGTAAACTAACTCGCAATAACACCT

282 Y Q L N P E O T E V L Y S E A V K A L D V D K E D H L I D A Y C G 314

1501 GTTGGACGATTGGATTTCCTTTGCAAGAAAGTAAAAACA CT CAGAGGTATGGATATTATTCCAGAAGCTATTGAAGATGCCAAGCGAAATGCTAAAA 1600  
CAACCTTGCTAACCTAAACGGAAACGTTTCCTTCATTTTGTGAGTCTCCATACCTATAATAAGGTCCTTCGATAACTTCTACGGTTTCGCTTACGATTTT

315 V G T I G F A F A K K V K T L R G M D I I P E A I E D A K R N A K R 348

1601 GAATGGGATTTGACAATCTCATTATGAAGCTGGAAACGGCAGAAGAGATTATTCCTCGTTGGTACAAGGAAGGCTACCGAGCAGATGCTTTGATTGTTCA 1700  
CTTACCTTAACTGTTATGAGTAATCTTCGACCTTGGCGTCTTCTCTAATAAGGAGCAACCATGTTCTTCGGATGGCTCGTCTACGAACTAACAACT

349 M C F D N T H Y E A G T A E E I I P R W Y K E G Y R A D A L I V D 381

1701 CCCACCAGTACAGGTCTGGATGATAAGTTATTAGATATCTTCTACTATGTACCAGAAAAATGGTTTATATTTCTTGTAAATGTTTCGACCTTCGGT 1800  
CGGTGGTGCATGTCCAGACCTACTATTCAATAATCTATGATAAGAAATGAATACATGCTCTTTTACCAATATAAAGAACATTACAAAGCTGGAAACCGA

382 P P R T G L D D K L L D T I L T Y V P E K H V Y I S C N V S T L A 414

1801 CGTGATTGGTACGCTTAGTAGAAGTCTATGATCTTCATTATATCCAGTCGGTGGATATGTTCCACATACAGCTCGAACTGAAGCTGTTGTAAAAATTAA 1900  
GCACTAAACCATGCGAATCATCTTCAGATACTAGAAGTAATATAGGTACGCCAGCTATACAAGGGTGATGTGAGCTTGACTTCGACAACATTTTAATT

415 R D L V R L V E V Y D L H Y I O S V D M F P H T A R T E A V V K L I 448

1901 TAACAAAAGTTTAAAAAGTAGTTGACAAAGTTTAAAAAGACTGTATAATAGTAAGAGTTGAAAAATAACAACTCAGGTTCGTTGGTCAAGGGGTTAAGAC 2000  
ATTGTTTTCAAAATTTTTCATCAACTGTTTCAAACTTTTCTGACATATTATCATCTCAACTTTTATTTGTTGAGTCCANGCAACCAAGTTTCCCAATTCTG

449 T K V • 452

2001 ACCGCTTTTCACGGCGGTAAACCGGTTGGAATCTCTACCGACTATGGTATGTTGCGGTTGGAACACTTGATGAAAAACTTTA 2084  
TGCGGAAAAGTGCCGCCATTGTGCCCAAGCTTAGGGCATGCTGATACCATAACAACGCCAACCTTGTAAGTACTTTTGAAT

gcp1315

Fig. 4

(SEQ ID NO:11) 1 AAGAGCTCCTTCCTTTTATTTATCTTAGCAAAATTCCTCAAATTAGCTAGTAGCATAGCCTGTTTGTACTGGCTAAAAACAGGCTATTTCAAATTCAG 100  
(SEQ ID NO:12) TTCTCGAGGAAGAAAAATAATAGAAATCGTTAAAGGGAGTTAAATCGATCATCGTATCGGACAAACATGACCGATTTTGTCCGATAAAGTTTAAAGTC

101 TTTCAGACCATCTAGCATGGAATAATCTGTTATAATAATGGAAAAGGAGAGCGCATGCACAGATTTTATTAATAGAAGATGATCAGGTCATTGCTCA 200  
AAAGTCGTAGATCGTACCTTTTAGACAAATATTATTACCTTTTCCTCTTCGGCTACGTTCTAAAAATAATTATCTTCTACTAGTCCAGTAAGCAGTT

(SEQ ID NO:10) 1 M H K I L L I E D D O V I R Q 15

201 CAGATTGGAAAAATGCTCTCAATGGGATTTTAAAGTGGTCTGTTAGAGACTTTTGAAGTTTTCAGTCTATTGTTCACTCGGAACCTCATCTGG 300  
GTCTAACCTTTTACGAGAGACTTACCCCTAAANTTACCAGGACCATCTTCTGAAATACCTTCAAACTCAGATAAAACAGTCAGCTTGGAGTAGACC

16 Q I G K M L S E W G F X V V L V E D F M E V L S L F V O S E P H L V 49

301 TCCTCATGGATATTGGTTTGGCTTGTATAATGGTTATCACTGGTGTGAGGAAATCGGCAAGATTTCCAAAGGTACCTATCATGTTTCTTTCTCGAGACA 400  
AGGAGTACCTATAACCAACGGGAACAAATTACCAATAGTGACCCACAGTCTTTAGGCGTTCTAAAGGTTCCATGGATAGTACAAAGAAAGAGCTCTCT

50 L M D I G L P L F N G Y H M C Q E I R K I S K V P I M F L S S R D 82

401 CCAGGCTATGGATATTGTCATGGCAATCAATATGGGGGCGGATGACTTTGTGACCAAGCCTTTTACCAGCAGGTTCTTTAGCTAAGGTTCAGGCGTTG 500  
GTCGAGTACCTATAACAGTACCGTTAGTTATACCCCGCTACTGAAACACTGTTTCGAAAACTGGTCGTCCAGAAAAATCGATTCCAAGTCCCGAAC

83 Q A M D I V N A I N M G A D D F V T K P F D O O V L L A K V O G L 115

501 TTGCGTCTTCTATGAGTTTGGGCTGATGAGAGTTTGTGGAATATGCTGGTGTATCTCAATACCAAAATCCATGGATTATCATTTCAAGGGCAAG 600  
AAGCGAGCAAGGATACCTCAACCCGCACTACTCTCAACGACCTTATACGACCACAAATAGGAGTTATGGTTTAGGTACCTAAATGTAATAGTTCCCGTTT

116 L R R S Y E F G R D E S L L E Y A G V I L N T K S M D L H Y Q G Q V 149

601 TCTTGAATTTGACCAAGAATGAATTCAGATTTTACGCGTGTATTGAGCATGAGGCAACATCGTAGCACGTGACGACCTGATGCGGAACTTTGGAA 700  
AGAACTTAACTGCTTCTACTTAAGGTCTAAATGCGCAATAACTCGTACGTCCGTTGTAGCATCGTGCACTGCTGGACTACGCGCTTGAACCTT

150 L N L T K N E F C I L R V L F E H A G N I V A R D D L M R E L M N 182

701 CAGTGACTTTTCATTGATGATAATACCTCTCTCAATGTGGCTCGTTTGGCTAAAAAGTTGGAGGAGCAGGGATTGGTAGGATTTATCGAGACCAAG 800  
GTCACTGAAAAAGTAACTACTATTATGGGAGAGACAGTTACCGGAGCAACGCATTTTCAACCTCTCGTCCCTAACCATCTAAATAGCTCTGGTTC

183 S D F F I D D N T L S V N V A R L R K K L E E Q G L V G F I E T K 215

801 AAAGCAATAGGGTACGGATTGAAGCATGCTTGAATGSAACAAATTTTCTAGCCTATCTGGCTCCCGTAGTCCTTTTATCTATCTGCTTTCTTC 900  
TTCTCTATCCCATGCTAACTTCGTACGAACTAACCTTTGTTAAAAAGATCGATAGACGCGAGCGCATCAGCAGAAAAATAGATAGACGAAAGAAAC

216 K G I G Y G L K H A \* 226

901 GCATTTCTTCTTACTCTTCAGTTTTATTGCCAGTCTAGGAATTAATCTCTACTTTTCTTCTTGTGTTGCTTTGTAAACCATCTTATTTTCA 1000  
CGTAAAGAACAGAAATGAGAAAGTCAAAATAAACCGTCAGATCTTAAATGAAGGAGATGAAAAAGAAGAACAAACGAAACATTGGTGAATAAAGT

gcp1493

Fig. 5

(SEQ ID NO:14) 1 TAAAGACACTGGAACGACCAACACCTTCCGATTTTAGGTAAAGAAAGCTGGTATGGCAACCTTTGTGATTGACTTTTCAAAGGAACCTTAGCAACCGTG 100  
(SEQ ID NO:15) 1 ATTTCTGTGACCTTGCTGCTGTGTGGAAGCGGTAAATCCATTCTTTGACCATACCGTTGGAACACTAACTGAAAAAGTTTCCTTGGGATCGTTGGCAG  
(SEQ ID NO:13) 1 K D T G T T N T F R I L G K K A G M A T F V I D F F K O T L A T L 33

101 CTTCCGATTATTTTCACTACAGGCGTTTCTCCTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTGCAGGATTTAAAGTG 200  
GAAAGGCTAATAAAAAGTAGATGTTCCGCAAGAGGAGAGTAGAAACCTGAAAGCCGACATAGCCGGTATGGAAGGGATAGAAACGTCCTAAATTTCCAC  
34 L P I I F R L Q G V S P L I F G L L A V I G H T F P I F A G F K G G 67

201 GTAAAGGCTGTCGCAACCACTGCTGGAGTGATTTTCGGATTTCGGCTATCTTCTGTCTCTACCTTGGGATTATCTTCTTTGGACTCTCATATCTTGGCAG 300  
CATTCGGACAGCGTTGGTCACGACCTCACTAAAGCCCTAAAGCCGATAGAAAGACAGAGATGGAACGCTAATAGAAAGAACTGAGAGTATAGAACCGTC  
68 K A V A T S A G V I F G F A P I F C L Y L A I I F F G L S Y L G S 100

301 TATGATTTCACTGTCTAGTGTACAGCATCGATCGCGCGCTGTTA 344  
ATACTAAAGTGACAGATCACAGTGTCTAGCTAGCGCGGACAT

101 M I S L S S V T A S I A A V 114

gcp1507

Fig. 6

(SEQ ID NO:17) 1 CTAAGGTAAATTGAATGAAAGTATAAAATTAATGCTCTATCTTACATGGGAATTCGTCTTGAATATTATTTTCCCATCCTAACTGGAACTATG 100  
(SEQ ID NO:18) GATTTCCATTTAACTTACTTTTCATATTTTAAATTTACGAGATAGAATGTACCCCTAAGCACAGAACTTATAATAAAAGGGTAGGATTGACCTTGGATAC  
(SEQ ID NO:16) 1 M K S I K L N A L S Y M G I R V L N I I F P I L T G T Y V 29

101 TCGCGCGTGTCTTGGACCGAACTGACTATGGTTACTTCAACTCAGTCGACACTATTTTGTCAITTTTCTTCCCTTTGCAACTTATCGTCTCTATAACTA 200  
AGCGCGCACAGAACTCGCTTGACTGATACCAATGAAGTTGAGTCAGCTGTGATAAAACAGTAAAAAGAACGGGAAACGTTGAATACCAACAGATATTGAT  
30 A R V L D R T D Y G Y F N S V D T I L S F F L P F A T Y G V Y N Y 62

201 CGGTTTAAGGGCTATCAGTAATGTCAAGGATAACAAAAAGATCTTAACAGAACCTTTCTAGTCTTTTATTTTGTGCGCTTGTACGATTTTGACC 300  
GCCAAATTCGGATAGTCATTACAGTTCTTATTGTTTTCTAGAATTTCTTGGAAAGATCAGAAAAATAAACACGTAGCGAACATGCTAAAACCTGG  
63 C L R A I S M V K D M R K D L N R T F S S L F Y L C I A C T I L T 95

301 ACTGCTGTCTATATCCTAGCCTATCCTCTCTTCTTACTGATAATCCAATCGTCAAAAAGGTCTACCTTGTATGGGGATTCAACTCATTGCCAGATT 400  
TGACGACAGATATAGGATCGGATAGGAGAGAAGAAATGACTATTAGGTTAGCAGTTTTCCAGATGGAAACAATACCCCTAAGTTGAGTAAACGGGTCTAAA  
96 T A V Y I L A Y P L F F T D N P I V K K V Y L V M G I O L I A Q I F 129

401 TTTCAATCGAATGGGTCAATGAAGCTCTGAAAATTACAGTTTCTCTTTTACAAAACCTGC 460  
AAAGTTAGCTTACCCAGTTACTTCGAGACCTTTAATGTCAAAGAGAAAATGTTTTGACC  
130 S I E W V N E A L E N Y S F S P T K L 148

9ep1511

Fig. 7

(SEQ ID NO: 20) 1 CCGCCATTTACCGTATGGATTTCACTATGTAATGATTTTATGGACAACTGAGAGCAGGACGAGGAAATGATGTTTGTGACGAGTTGCTATACA 100  
(SEQ ID NO: 21) CGAGCGTAAATGGCACTACCTAAAGTGCAATCACTACTAAAAATACCTGTTGCAGCTCTGCTCCTGCTCTTACATACAAAACACTGCTCAACGATATGT

101 GGGAGTAGGCATGCAGATTCAAAAAAGTTTTAAAGGGCAGTCTCCCTATGGCAAGCTGTATCTAGTGGCAACGCCGATTGGCAATCTAGATGATGACT 200  
CCCTCATCGGTACGCTAAAGTTTTTCAAAATTTCCCGTCAGAGGGATACCGTTGACATAGATCACCGTTGCGGCTAACCGTTAGATCTACTATACTGA

(SEQ ID NO: 19) 1 M Q I O K S F K G O S P Y G K L Y L V A T P I G N L D D M T 30

201 TTTGCTGCTATCCAGACCTTGAAGAAGTGGACTGGATTGCTGCTGAGGATACGCGCAATACAGGGCTTTTGTCTCAAGCATTTTGACATTTCCACCAAGC 300  
AAAGCAGATAGGTCGGAACCTTTCTTCACTGACCTAACGACGACTCTATGCGGCTATGTCCGAAACGAGTTGTAAGGTTGCTTTCG

31 F R A I O T L K E V D W I A A E D T R N T G L L L K H F D I S T K O 64

301 AGATCAGTTTTCATGACCAATGCAAGGAAAAAATTCCTGATTTCATTGGTTTCTGAAAGCAGGGCAAGTATTGCTCAGGTCTCTGATGCCGCTTT 400  
TCTAGTCAAAAGTACTGCTTACGTTTCTTTTAAAGGACTAACTAACCAAGAACTTTCTCCGTTTCATAAGGAGTCCAGAGACTACCGCCAAA

65 I S F H E H N A K E K I P D L I G F L K A G O S I A O V S D A G L 97

401 GCGTAGCATTTTCAAGCCCTGCTCATGATTAGTTAAGGCAGCTATTGAGGAAGAAATTCAGTTGTGACTGTTCCAGGTACCTCTGACGGAATTTCTGCC 500  
CGGATCGTAAAGTCTGGGACGAGTACTAAATCAATTCCTCGATAACTCTCTTTAAAGTCAACACTGACAAAGTCCATGGAGACGCTCTTAAAGACGG

98 P S I S D P G H D L V K A A I E E E I A V V T V P G T S A G I S A 130

501 TTGATTGCCAGTGGTTTACGCCACAGCCACATATCTTTTACGTTTTTTACCGAGAAATCAGGTCAACAGAAAGCAATTTTGTGCTCAAAAAAGATT 600  
AACTAACGGTCACCAATCGCGGTGTCGGTGTATAGAAAATGCCAAAAATGGCTCTTTAGTCCAGTTGTCTTGGTAAAAAACCGAGATTTTCTAA

132 L : A S G L A P O P H I F Y G F L P R K S G O Q K O F P G S K K D Y 164

601 ATCGTGAACACAGATTTTATGAATCACTCATCGGTAGCAGACAGTTGGAAAAATATGTTAGAAGTCTACGGTGACCGCTCGGTTGTTTGGTCAAG 700  
TAGGACTTTGCTCTAAAAATACTTAGTGGAGTAGCACATCGTCTGTGCAACCTTTTATACATCTTCAGATGCCACTGGCGAGCCAAACAAACAGTCTC

165 P E T Q I F Y E S P H R V A D T L E N H L E V Y G D R S V V L V R 197

701 GGAATTGACCAAAATCTATGAAGAATACCAAGAGGTACAAATTTCTGAATTCGTGGAAGCATCTCTGAAACGTCTCTCAAGGGTGAATGTCTTCTGATT 800  
CTTTAACTGGTTTATGATCTTTATGGTTTCTCATGTTAAAGACTTAACGACCTTTGTAAGAGACTTTGAGAGAGTTTCCCACTTACAGAGAGCTAA

198 E L T K I Y E E Y O R G T : S E L L E S I S E T S L K G E C L L I 230

801 GTTGAAGGTGCCAGCAAGGTGTGGAGGAAAGGATGAGGAAGACTTGTCTTAGAAATCCAGCCCGTATCCAGCAAGGCATGAAGAAAAATCAAGCTA 900  
CAACTTCCAGCGTCTTTCCACACCTCTTTTCTACTGCTTCTGAACAAGAACTTTAGGTTCCGGCATAGGTGTTCCGTACTTCTTTTATGTTGGAT

231 V E G A S K G V E E K D E E D L F L E I Q A R I Q Q G M K K N Q A : 264

901 TTAAGGAAATAGCTAAGATTTACCAAGTGAATAAGAGTCAACTCTACGCTGCTTACCAGCACTGGGAAGAAAAACAATAAGGGAGACAGGATGTATAAA 1000  
AATTCCTTTATCGATTCTAAATGGTCACTTATCTCAGTTGAGATGCGAGGATGGTCTGACCTTCTTTTGTATTTCCTCTGTCTACATTATT

265 K E I A K I Y O W N K S C L Y A A Y H D W E E K Q \* 290

g9p1518

Fig. 8 (Sheet 1 of 2)

(SEQ ID NO: 23) 1 ATGCGCTTGGTTAAAAAAGGTGGCAATGCTCTTTAAGTGCAAGTTATTGCGCTGTAGCATATAAATCTATTTCTACATATTTTAAAAAGTTCTAGGAC 100  
(SEQ ID NO: 24) TACCGAACCAATTTTTCACCGTTAGCAGAAATTCAGGTTCAATAACCGACATCGTATATTAGATAAAGGATGTATAAAAAATTTGCAAGATGCTC

201 TTAAATTTGAAACGTTTAGCTTGTGGTATAATAGATTTATGGATAAAAAATATGAAAAATCTCTCAGGATTTGGGAGTGACGTTAAAGCAAAATTGATACC 200  
AATTAACCTTTGCAAAATCGAACACCATATTATCTAAATACCTATTTTATACCTTTTATAGAGAGTCTTAAACCTCACTGCAATTTGTTTAACTATCG

(SEQ ID NO: 22) 1 M D K K Y E K I S O D L G V T L K Q I D T 21

201 GTTCTAAGTTTGACAGCTGAAGGGGCGACTATTCCTTTATCGCGCTTATCGCAAGGACATGACTGGTAGTCTGGATGAGGTGGCGATTAAAGGCTATTA 300  
CAAGATTCAAACCTTCGACTTCCCGCTGATAAGGGAAATAGCGGCAATAGCGTTCTGTACTGACCATCAGACCTTACTCCACCGCTAATTCGATAAT

22 V L S L T A E G A T I P F I A R Y R K D M T G S L D E V A I K A I I 55

301 TTGATTTGATATAAAGTCTGACAAATCTCAATGACCGTAAGCAAGCTGTCTTAGCTAAGATTCAAGAACAAAGGTAAGTTGACCAAGGAATTGCAAGAAC 400  
AACTAAACCTATTTTACAGACTGTCTTAGAGTTACTGGCATTCCTTCGACAGAAATCGATTCTAAGTTCTTGTTCATTCAACTGGTTCTTAACTTCTTTCG

56 D L D K S L T N L N D R K E A V L A K I O E Q G K L T K E L E E A 88

401 TATCTTAGTTGCGGAAAAATTAGCAGACGTTGAAGAACTCTATCTTCTTATAAGGAAAAAGCGTCGTACCAAGGGCAACCTATTCCTGCTGAAGCTGGA 500  
ATAGAAATCAACGGCTTTTAAATCGTCTGCAACTCTTGAGATAGAAAGGAATATTCCTTTTCGAGCATGGTTCCGTTGGTAAACGGGCACCTTCGACCTGAG

89 I L V A E K L A D V E E L Y L P Y K E K R R T K A T I A R E A G L 121

501 TTTCTCTTCTGCTGCTTTGATTTGACAAATATAGTTGACTTAGAGAAAGAAAGCTGAAAGTTCTGCTGTGAAGGATTTGGAGCTGGCAAGGAAGCCCTTGA 600  
AAAGGAGAACGAGCAAACTAAACGCTTTATATCAACTGAATCTCTTCTTCGACTTTTCAAGCAGACACTTCCTAAACGCTGAGCCTTCTTTCGGAACT

122 F P L A R L I L Q N I V D L E K E A E K F V C E G F A T G K E A L T 155

601 CCGGTGCAGTTGATATTTGCTGCAAGCCTTATCGGAAGATGTGACCTTGGCTCTATGACTTATCAGGAAGTCTGAGACACTCTAAACTCACTTCTCA 700  
GGCCAGCTCAACTATAAAACAGCTTCGGAATAGCCTTCTACACTGGAACGCAAGATACTGAATAGTCTTCAAGACTCTGTGAGATTGAGTGAAGAGT

156 G A V D I L V E A L S E D V T L R S M T Y Q E V L R H S K L T S O 188

701 AGCCAAAGGATGAAAGCTTGTGATGAAAAGCAGCTTTTCAGATTATTATGATTTTCAGAGACAGTTGGAACTATGCAAGGCTATCGTACCTTGGCTCTC 800  
TCGGTTCCTACTTTTCAAGAACTACTTTCTGTCGAAAAAGTCTAAATAATCTAAAAAGTCTCTGTCAACCTTGATAGCTTCGATAGCATGGAACCGAGAG

189 A K D E S L D E K O V F Q : Y Y D F S E T V G T M Q G Y R T L A L 221

801 AATCGTGGGAGAAAGCTTGGTGTCTTGAAGATCGCTTTGAACATCGGACGGACCGTATTTCTGCTCTCTTGTACTCTTTCAAGGTGAAAAATGCTT 900  
TTAGCACTCTCTTTGAACCAAGAACTTCTAGCCAAAATCTGTACGCTGCTTGGCATAAGAACGGAAGAAAGATGAGCAAGTTTCACTTTTACGAA

222 N R G E K L G V L K I G F E H A T D R : L A F F A T R F K V K N A Y 255

921 ATATTGATGAAGTTGTTTACGCAATCGTTAAGAAAAAGGCTTGGCTGCTATTGAGCGTCTGATTTCGACAGAAATTAAGTGAAGAGCTGAAGAGGGAGC 1000  
TATAACTACTTCAACAAAGTCTTAGCGAATTTCTTTTCAGAACGAGGATAACTCGCAGCATAGCGCTGTCTTAATTGACTCTTTCAGCTTCTCCCTCG

256 : D E V V O Q S V K K K V L P A I E R R I R T E L T E K A E E G A 288

1001 TATCCAATTTTCTGACAAATCTGCGCAATCTCTTGGTGTCTCACTGAAAGGGCGCGTGGTTCTTGGATTTGACCCAGCCTTTCTGACAGGTGCC 1100  
ATAGGTTGAAAAAAGACTGTTAGACGCGTTAGAGGAGAACCAACGAGGTGACTTCCCGCGCACCAAGAACTTAACTGGGTGGGAAAGCATGTCCACGG

289 : Q L F S D N L R N L L L V A P L K G R V V L G F D P A F R T G A 321

1201 AAGTTAGCTGTCTGATGCAACAGGAAAAATGCTGACAACTCAGGTTATTATCTGTTAAACCAAGCATCAGCTCTTCAATCGAAGAACCCAGAAAG 1200  
TTCAATCGACAGCACCTACGTTGTCTTTTACGACTGTTGAGTCCAATAAATAGGACAAATTTGGTCTGAGTGGAGCAGTTTACGCTTCTTGGTTCTTTC

322 K L A V V D A T G K H L T T O V : Y P V K P A S A R Q I E E A K K D 355

1301 ATTTAGCAGATTTAATTGCTCAATACGGTGTAGAGATTATTGCCATTGCGAAATGGAACGCGCAAGTCTGTAAGGTGAAGCTTTGTAGCGCAAGTTCTGAA 1300  
TAAATCGTCAAAATTAAACAGTTATGCCACATCTCTAATAACGGTAACTTTACCTTGGCGGTCAGCACTTTCACCTCGAAGAACCTCGCTTCAAGACTT

356 L A D L I G C Y G V E : I A I G N G T A S R E S E A F V A E V L K 388

Fig. 6 (Sheet 2 of 2)

10/30

1301 AGATTTCCTGAAGTCAGCTATGTTATCGTTAATGAAAGTGGTCTCTGTCTATTCTGCCAGCGAACTTGCTCGTCAGGAGTTTCCAGACTTGACCGTT 1400  
TCTAAAGGGACTTCAGTCGATACAATAGCAATTACTTTCAACACGAAGACAGATAAGAAGTCTGCTTGAACGAGCAGTCTCTCAAAGGTCGTGAATGGCAA

389 D P P E V S Y V I V N E S G A S V Y S A S E L A R O E F P D L T V 421

1401 GAAAAACGCTCTGCCATTTCTATCGCCCTCGTTTGCAGATCCTCTTGGGAATTGGTCAAAATCGATCCTAAGTCAATTGGTGTGGTCAATACCAAC 1500  
CTTTTTCGAGACCGTAAAGATAGCGGGCAGCAACGTTCTAGGAGAACGCTTAACCAAGTTTATAGCTAGGATTCAGTTAAACACAGCCAGTTATGGTTG

422 E K R S A I S I A R R L O D P L A E L V K I D P K S I C V G Q Y Q H 455

1501 ACCATGTCAGTCAGAGAAACTATCTCAGAGTCTCGACTTTGTTGTGATACAGTGGTTAACCAAGTTGGTGTCAATGTCAATACAGCTAGCCAGCTCT 1600  
TGCTACAGTCAGTCTTCTTTGATAGACTCTCAGACCTGAAACACAGCTATGTACCAATTGGTTCACCAAGTTACAGTTATGTGATCGGTCGAGA

456 D V S Q K K L S E S L D F V V D T V V N Q V G V N V N T A S P A L 488

1601 TCTTTCAACGTAGCTGGACTCAACAAAATATCTCTGAAAATATTGTCAAATACCGCGAGGAAGAAGGAAAAATCACTTCAACGCGCCCAATCAAGAAA 1700  
AGAAAGTGTGCATCGACCTGAGTTGTTTGTATAGAGACTTTTATAACAGTTTATGGCGCTCCTTCTTCTTTTATGTAAGTCCGCGGCTTATGTTCTTT

489 L S H V A G L N K T I S E N I V K Y R E E E G K I T S R A Q I K K 521

1701 GTTCTCTGTCGGGAGCCAAAGGCTTTGAGCAGGCTGCTGGTTCTCTTCGTATCCTTGAAAGTAGCAATATCCTTGATAATACAGGAGTTCACCCAGAG 1799  
CAAGGAGCAGACCTCGGTTCCGGAACTCGTCCGACGACCAAGGAAGCATAGGCACTTTCATCGTTATAGCAACTATTATGTCTCAAGTGGGTCTC

522 V P R L G A K A F E Q A A C F L R I P E S S N I L D N T G V H P E 554



gepl546

Fig. 9

11/30

(SEQ ID NO: 26) 1 TACTGGGGCAAGGGTTTCTTACCTGTTCTGAATGTGAAGGTCTTTCTTGAAAATGGTGAAGTTAAGATTTTCAGAGCACTCAACGAAGCCAGNATCCGC 100  
(SEQ ID NO: 27) ATGACCCCGTTCCCAAGAATGGGACAAGACTTACACTTCCAGAAAGAACTTTACCACTTCAATTCATAAAGTCTCGTGAAGTTGCTTCGGTCTATAGCGG  
(SEQ ID NO: 25) 1 T G A R V S Y P V L N V R V F L E N G E V K I P R A L N E A X I R 33

101 AGGTCTGATCGAACCATGGTGGCAGATATTGTAAATAAGGTGTTCCCTTTGAACGTTTTCTGCAGACGGCTAAACAGTTTCCACACCGACTGGTAGTA 200  
TCCAGACTAGCTTGGTACCACCGTCTATAACATTATTTACCACAGGGGAACTTGCAAAAGCACCTCTGCCCGATTGTCAAAGCTGTGGCTGACCATCAT  
34 R S D R T M V A D I V I N G V P P E R F R G D G L T V S T P T G S T 67

201 CTGCCTATAACAAGTCTCTTGGCGGTGCTGTTTACACCTTACCATTGAAGCTTTGCAATTAACGGAGATTGCCAGCCTTAATAATCGTGTCTATCGAAC 300  
GACGGATATTGTTTACAGAAACCGCCACGACAAAATGTGGATGGTAACCTTCAAAACGTTAATTGCTCTAACGGTCCGAATTAATTAGCACAGATAGCTTG  
68 A Y N K S L G G A V L H P T I E A L O L T E I A S L N M R V Y R T 100

301 ATTGGGCTCTTCCATTATTGTGCCTAAGAAGGATAAGATTGAACCTTATTCACAAAGAAACGATTATCATACTATTTGGTTGACAATAGCGTTTATTCT 400  
TAACCCGAGAAGGTAAATAACACGGATTCTTCTATTCTAACTTGAATAAGGTTGTTCTTTGCTAATAGTATGATAAAGCCAACTGTTATCGCAATAAGA  
101 L G S S I I V P K K D K I E L I P T R N D Y H T I S V D N S V Y S 133

401 TTCGTAAATTTGAGCGTATTGAGTATCAAAATCGACCATCATAGATTCACTTTGTGGGACTCTAGCCATACCACTTCTCGAACCGGTGTTAAGGATC 500  
AAGGCATTATAACTCGCATAACTCATAGTTTAGCTGCTAGTATTCTAAGTGAACACGGCTGAGGATCGGTATGGTCAAAGACCTTGGCAAAATTCCTAC  
134 F R N I E R I E Y Q I D H H K I H F V A T P S H T S F W N R V K D A 167

501 CCTTTATCGGTGAGGTGGATGAATGAGGTTTGAATTTATCGCAGATGAACATGTCAAGGTTAAGACCTTTTAAAAAA 578  
GGAAATAGCCACTCCACCTACTTACTCCAACTTAAATAGCGTCTACTGTACAGTTCCAATTCGGAAAAATTTTTT  
168 F I G E V D E \* 175

gcp1551 Fig. 10

(SEQ ID NO: 29)<sub>1</sub> CGCTCTAAAAGAAACCTACTGCGAGCTGATAGATGGGAAGTACTATTATTTTGATCCTTTATCCCGAGAGATGGTTGTGGGCTGGCAATATATACCTGCT 100  
(SEQ ID NO: 30) CCGAGATTTTCTTTGGATGACCTCTCACTATCTACCTTTCATGATAATAAACTAGGAAATAGGCCTCTCTACCAACAGCCGACCGTTATATATGGACCA  
(SEQ ID NO: 28)<sub>1</sub> M V V G M Q Y I P A 10

101 CCACACAAGGGGTTACGATTGGTCCTTCTCCAAGAATAGAGATTGCTTTAGACCAGATTGGTTTATTTTGGTCAAGATGGTCTTACAAGATTGG 200  
GGTGTGTTCCCCAATGCTAACCGAAGAGGTTCTTATCTTAACGAGAATCTGGTCTAACCAAAATAAAACCAAGTTCTACCAAGAAATGTTCTTAAAC

11 P H K G V T I G P S P R I E I A L R P D W F Y F G Q D G V L O E F V 44

201 TTGGCAAGCAAGTTTGAAGCAAAAACTGCTACGAATACCAACAAACATCATGGGAAGAAATATGATAGCCAAGCAGAGAAACGAGTCTATTATTTGA 300  
AACCGTTGGTTCAAAATCTTGGTTTTGACGATGCTTATGGTTGTTTGTAGTACCCCTTCTTATACTATCGGTTGGTCTCTTTGCTCAGATAATAAACT

45 G K O V L E A K T A T N T N K H H G E E Y D S Q A E K R V Y Y F E 77

301 AGATCAGCGTAGTTATCATACTTTAAAACTGGTTGGATTATGAAGAGGGTTATTGGTATTATTTACAGAAGGATGGTGGCTTTGATTCTGGCATCAAC 400  
TCTAGTGGCATCAATAGTATGAAATTTTGACCAACCTAAATACTTCTCCAAATAACCATATAAATGTCTTCTACCAACCGAAATGAAGAGCGTAGTTG

78 D Q R S Y H T L K T G W I Y E E G Y W Y Y L Q K D G G F D S R I N 110

401 AGATTGACGGTTGCGAGCTAGCAGTGGTTGGGTTAAGGATTACCCCTCTTACGTATGATGAAGAGAAGCTAAAAGCAGCTCCATGGTACTATCTAGATC 500  
TCTAACTGCCAACCTCTCGATCGTGCACCAACCAATTCCTAATGGGAGAATGCATACTACTTCTCTTGGATTTCGTGGAGGTACCATGATAGATCTAG

111 R L T V G E L A R G W V K D Y P L T Y D E E K L K A A P W Y Y L D P 144

501 CAGCAACTGGCTGGCAAAACCTTGGGAACAAATGGTACTACCTCCGTTTATCAGGAGCTATGGTAACTGGCTGGTATCAAGATGGTTTAACTGGTACTA 600  
GTGGTTGACCGACCGTTTGGAAACCTTGTTTACCATGATGGAGGCAAGTAGTCTCGATACCATTGACCGACCATAGTTCTACCAAAATTGAACCATGAT

145 A T G M Q N L G N K W Y Y L R S S G A M V T G W Y Q D G L T W Y Y 177

601 CCTAAATCGAGTAATCGAGACATGAAGACAGGTTGGTCCAGTCAATGGTAAGTCTATGCTATGCTATGATTGAGGTGCTTTAGCTGTTAATACCACA 700  
GGATTTAGTCCATTACCTCTGTAATCTGTCCTCAACCAAGGTTCACTTACCATGACCATGATACGGATACTAAGTCCAGGAAATCGACAAATTATGGTGT

178 L N A G N G D M K T G W F Q V N G N W Y Y A Y D S G A L A V N T T 210

701 GTAGGTGGTTACTACTTAACTATATGGTGAATGGGTTAAGTAATGAAGGCTAATTGTAACTGTGATGGATACTTAACTTTGTATAATAGGTGGATAA 800  
CATCCACCAATGATGAATTTGATATTACCACTTACCAATTCACTTCTCGATTAACTTTGACACTACCTATGAATTGAAACATATTATCCACCTATT

211 V G G Y Y L N Y N G E W V K \* 225

gcp1561

Fig. 11

(SEQ ID NO: 32) 1 TTTTATGGATATTTATTAAGAAGCCATTATTCACCACTTCAGTCCGATGATACCGAGCTGTTCTTAGCAGATAAGTTTCTCAATATTACTCCAAA 100  
(SEQ ID NO: 33) AAAATACCTATAAATATAATTCTTTCCGTAATAAGTGGTCAAGTCAAGCCTACTATGGCTCGACAAGAAATCCTCTATTCAAAGAGTTATAATGAGGTTTT  
(SEQ ID NO: 31) 1 M D I Y I R K A I I H O F S P D D T E L F L A D K F L N I T P K 32

101 ATCGAAGAATACCTACGTAAAAAATGGAACATGTGTATT CAGATGAAGCCAGACTGGGATTTCGAAGAAGAAAAATCCCTTCTTCAATCATATTACAG 200  
TAGCTTCTTATGGATGCAATTTTAACTTGTAACATAAGCTACTTCGGTCTGACCTAAAAGCTTCTTCTTTTAGCGAAGAGTTAGTATAATGTC  
33 I E E Y L R K K I E H V Y S D E A K T G I F E E E N P F F N H I T D 66

201 ACGATTGTTGGAGACATCAGTAACGCTGGCTAATCTCTGAAAGAGGAGTTTAGCATTCTGAAAAATCTCAAGACCAATGACTTGATTGTTTCAATT 300  
TGCTAAACAACCTCTGATGTCATTGGCAGCGATTAGAGACCTTCTCTCAATCGTAAGACTTTTAGAGTTCTGGTTACTGAACTAAAAACAAGTTAA  
67 D L L E T S V T L A N L W K E E F S I S E N L K T N D L I F V Q F 99

301 TTCTAAAGAAGGTGTAGAACAATTCGCTTCTTGGCAATTCGCTGCGGAGAGCTTGACCCACCTCGAGGAGAGAGTTGATAATCCAATCAAGCTGACT 400  
AAGATTCTTCCACATCTTGTAAAGCGAAAGAACGCTTAACGGGACGCGCTCTGGAACCTGGGTGGAGCCTCTCTTCAACTATTAGGTTAGTTGCACTGA  
100 S K E G V E H F A F L R I A L R E T L T H L G G E V D N P I K L T 132

401 CAGAATAACCTGCTCGATTTCGAACCGGTGCTGACGAGGCTTGGTGGTCAATCTTCAGAGTCGCAAGTATCACCTGATTGAAAAACGAATCAAGTACA 500  
GTCTTATGGACGGACCTAAACCTTGCCACGACTGCTCGGAACCAACAGTATAGAAGTCTCAGCGTTTCATAGTGGACTAACTTTTCTGCTAGTGTCTG  
133 Q N N L P G F G T G A D E A L V V N L O S R K Y K L I E K R I K Y N 166

501 ACGGGACTTTTGAACATTTTTCAGATAATCTTCTGCTGCTGCTTCAAGATTCTCTTAAAAATCTATCAAGGAATCGGAAAAACAGCCAGAG 600  
TGCCCTGAAAAAATCTGATAAAAGTCTATTAGAAGAACGACAGCGAGGATCTTAAGAGGATTTTATAGATAGTTCTTGAACCTTTTGTGGGTCTC  
167 C T F L N Y F S D N L L A V A P K I S P K K S I K E L E K T A Q R 199

601 AAATGGCTGAATCTTTTAAACACAGATGATTTCAATTTCAATCCAAGCTCAATCAGCTATTTTCAACCACTAGAGAAGCAATGAATGTCACCTGAG 700  
TTAACCGACTTAGAAAAATGTGTCTACTAAAAAGTTAAAGTTAGGTTCCAGTTTAGTCGATAAAAGTTGTTGGATCTTCTTCTGTTACTTAACAGTGGACTC  
200 : A E S F N T D D F O F O S K V K S A I F N N L E E S N E L S P E 232

701 AAATGGCTAATGACCTTTTGAACAATCTGACGGCTCGTTTGAAGCTTTATTGACCAAGTCAGAGAAGCCGTACCAGAACCTGTTCAATTTGATGAAA 800  
TTAACCGATTACTGGAAAAATGTTGTTAGACTGCGAGCAAACTCGAAATAACTGGTTCAAGTCTCTTGGCATGGTCTTGAGCAAGTTAAACTACTTT  
233 K L A N D L F D N N L T A R L S F : D Q V R E A V P E P V Q F D E I 266

801 TCGATGCCAGTCGCCAATTAAGAAAAATGAAACCAAAACTCTCTTATCAATGGAAATGAGCTCATCGTTCCCAATTAACCTCTATCAAGACGCCGA 900  
AACTACGGTCAGCGGTAAATTTCTTTAACTTTTGGTTTTGAGAGGAATAGTTTACCTTAACTCGAGTAGCAAGGGTTATTGAGATAGTTCTGCGGCT  
267 C A S R Q L K K F E N Q K L S L S N G I E L I V P N N V Y Q D A E 299

901 GTCTGTTGAGTTTATCCAAAACGAAATGGAACCTACTCTATCTTAATCAAAATATCGAGGATATCCAAAGTAAATAATGTTTAAACGAATTCGAAGAG 1000  
CAGACAACTCAATAGGTTTTCCTTTTACCTTGGATGAGATAGAAATAGT:TTTATAGCTCTATAGGTTTCATTTATTAGAAATTTGCTTAAGCTTCTC  
300 S V E F : Q N E N G T Y S I L : K N I E D I O S K \* 325

1001 TGGTGTACTAGCAGTCTTCTTTTCTGGCTATAAAGCTTACCGGTTTATCAAGATGTCAAAACAGTCATGACCTATCAACCCATGGTGGAGAAAT 1100  
ACGAACATGATCTCAGAAGGAAAAACGACCGATTTTCAATGGCGCAAGTAGTTCTACAGTTTCTTCAGTACTGGATAGTTGGGTACCAAGCTCTTTA

gcp1580

Fig. 12

(SEQ ID NO: 35) 1 AAATGTGCTATAATACTACAAAAAATCTTGTGAGGTTCCATATGCGAATATTTTTCATGATTTTCTGATTGTTTGTGTGCTCTATTGGTGATAGTC 100  
(SEQ ID NO: 36) TTTACACGATATTATGATCTTTTATGAACACCTCCAGGTAAATACGTTATAAAAAGTACTAAAAAGACTAACAAACACACGAGGATAACCACTATCAG  
(SEQ ID NO: 34) 1 M A I F F M I F L I V C V L L L V I V 19

101 ACACTGAGTACAGTTTATGTGCTTCTGAGCAGTCGGTGGCGATTATTGAACGCTTTGGGAAATACCAAAAGGTTGCTAATAGCGGTATTATATTGGCT 200  
TGTGACTCATGTCAAATACACCAAGCAGTCGTGAGCCACGCTAATAACTTGGGAAACCCCTTATGGTTTCCAAACGATTATCGCCATAAGTATAAGCGA  
20 T L S T V Y V V R Q Q S V A I I E R F G K Y O K V A N S G I H I R L 53

101 TGCCTTTTGGGATTGACTCGATTGCAGCAGGATTGAGTGGCTTGTTCGAAAGTGATATTGTTGTTGAGACTAAGACCAAGGACAAATGTTTCTGTTAT 300  
ACGGAAACCCCTAAGTAACTGAGCTAAGCTGCTGCTAAGTCAACGGCAACAGCTTTCACTATAACACCAACTCTGATTCTGGTTCTGTTACACAGCAATA  
54 P F G I D S I A A R I O L R L L O S D I V V E T K T K D N V F V M 86

101 GATCAATGTAGCGACTCAGTACCGTGTCAACGAGCAGAGCGTCAAGATGCTTACTATAAATCTATAGCTCCAGAACTCAGATTAATCTTATATCGAA 400  
CTACTTACATCGCTGAGTCATGGCAGTGTGCTGCTGCTGAGTGTCTACCAATGATATTGAGTATGCGAGTCTTAGAGTCTAATTTAGAAATATAGCTT  
87 M N V A T Q Y R V N E Q S V T D A Y Y K L I R P E S O I K S Y I E 119

101 GATGCTCTTCGCTCTTCTGTTCCAAAATTAACTTGGATGAATTGTTTGAGAAAAAGATGAGATTGCGCTTGAGGTTCAACACCAAGTAGCAGAAGAA 500  
CTACGAGAAGCGAGAAGACAAGGTTTAAATGGAACTACTTAAACAACTCTTTTCTACTCTAACGGGAACCTCAAGTTGTTGTTTATCGTCTCTTTT  
120 D A L R S S V P K L T L D E L F E K K D E I A L E V O H O V A E E M 153

101 TGACCACTTACGGCTACATTATCGTGAAACCTTGATTACCAAGGTGAAACGAGATGCGAGAAGTTAAGCAATCTATGAATGAAATCAATGCGCGCAACG 600  
ACTGGTGAATGCGGATGTAAAGCACTTTTGGAACTAATGTTCCAGCTTGGTCTACGTCCTCAATTGTTAGATACCTTACTTTAGTTACGCGCGGTTGC  
154 T T Y G Y I I V K T L I T K V E P D A E V K Q S M N E I N A A O R 186

101 TAACCGGCTCCGAGCAAGAAATCCGCGAAGCTGACAACTTAAATTTGCTGCTGAGCTGAAGCGGAGCAGAAAAAGACCGCTTCTATGTTGTTGGG 700  
ATTGCGCCAGCGTCTGTTCTTAAACCGCTTCTGAGCTTCTAATTTTAAAGTGAAGTCTGAGCTTCCGCTTCTGCTTTTCTGCGGAAAGTACCAACCCCT  
187 K R V A A Q E L A E A D X I K I V T A A E A E A E K D R L H G V G 219

101 ATTGCCCAACCAAGCTAAGGCGATTGTGGATGGATTGGCAGAGTCTATCAGCAACTCAAGGAAGCCAATGTTGGCATGACAGAAGAACAAATCATGTCTA 800  
TAACGGGTTGTTGCTTCCGCTAACCACTACCTAACCGTCTCAGATAGTGGCTTGAGTTCTTCCGTTTACCAACCGTACTGTTCTTCTGTTTATGACAGAT  
220 I A Q O R K A I V D G L A E S I T E L K E A N V G M T E E Q I M S I 253

101 TCTCTTTGACCAACCACTATTGATACCTTGAATACCTTTGCTCTAAAGGAAATCAAAACCTCTTTTACCAAAATCTCCAAATGGTGTGGATGATAT 900  
AGGAGAACTGGTGGTCTAATAACCTATGGAACCTATGGAACCGAGATTTCTTTAGTTTGGTAGAAAAATGGTTTATGAGGTTTACCACACCTACTATA  
254 L L T N O Y L D T L N T F A S K G N O T I F L P N T P N G V D D I 286

101 CCGTACACAAATCTTGTGAGGCTTCCGCGTGAAGAAATAATAGACTAATACTCTTGGAAATCTCTTCAAACTAGCTCAGCGTCTGTTCCCGTATA 1000  
GGCATGTGTTAGAACAGTCCGGAGCGGAGCTCTCTTATTTATCTGATTATGAGAAGCTTTAGAGAAGTTTGATGCGCTGCGAGCAGAACGGCATAT  
287 P T O I L S A L R A E K K \* 300

gsep1713

Fig. 13

(SEQ ID NO: 38) 1 CCTTCATATCGTGGATAAAATAGGGTTTTATTTGGAAAACTTTCTTTGTNTTCAAAATGCTAAAAAATGGTACAATANAGGAAAGCTTACTATTA 100  
(SEQ ID NO: 39) GGAACATATACCACTATTTATCCCAAAATAAAACTTTGCAAGGAAACAAAGTTTAAACGATTTTTHACCATGTTATNTCTTTGCAATGATAAT

101 TCTGAATCAGCAGATTTGAGAGAAAGGATTCATTTTGAATCAATAGGCTTTATTCAAAAGCTCAAGGGGTTGTCTAGTAAGAGCTGATTTTATTGGG 200  
AGACTTAGTCTGTAAACCTCTCTTCTTAAGTAAAACTTTAGTTATCCGAAATACTTTTCAGCTTCCCCAACAGATCATTTCTGACTAAAAATAACCC

(SEQ ID NO: 37) 1 L K S I G P I E K L K G L S S K E L I L L G 22

201 AATTATCCTAAGTATCTTTTACCCCTTTATCTTTTGTAGTTTACTCTGTTTATATATTATCAGTTTGTATTTTACAGGAGACATGAAAAGTATTCTT 300  
TTAATAGGATTCATAGAAAAATGGGAAATAGAAAAACATCAACATGAGACAAATATATAATAGTCAAACTAAAAATGTCTCTGTACTTTTCATAAGAA

23 I I L S I F L P F Y L F V V V L C L Y I I S L I F T G D M K S I L 55

301 CAGAAAAATGGGGAGCATCCGATGCTCTCTTTCTTAGCTATAGTACTGTTTATCCATCTTTCACAAAAATGGAGCGGCTCTTGTGGCTTCAGTAG 400  
GTCTTTTACCCCTCTGATGGCTACGACGAGAAAAAGAAATCGATATCATGACAAATATAGGTAAAGAACGTGTTTAACTACCCAGAACACCGAAGTCATC

56 Q K M G E H P M L L L F L S Y S T V I S I L A Q N W M G L V A S V G 89

401 GAAATGTTCTATTACTATTTCTTTTGCACATCAGTCGATTTTATCCATAAAATCTTTCGATTGATTTTTCAGTTCTGTCTTGTAGTGTCTT 500  
CTTACAAGATAAATGATAAAAGAAAAAGCTGATAGTCAGCTAAAAATAGGTTATTTAAGAAAGCTAACTAAAACTCAAGCAGAACAAACCATCACAGAA

90 M F L F T I F F L H Y Q S I L S H K P F R L I L O F V L F G S V L 122

501 GTCAGCTGCTTTTGCAGTTAGAACATTTCCAAATGCGAAGAAATTAACATGCTTTCTTTCACCAATATGCAGGTGTGGCATCAGAACCGGGCA 600  
CAGTCGACGAAAAACGGTCAAACTCTGTAAGGTTTAACTACTCTTAAATGATACGAAAAAGAAAGTGGGTTATACGTCACACCGGTAGTCTTGGCCCGT

123 S A A F A S L E H F G I V K K F N Y A F L S P N M Q V W H Q N R A 155

601 CAAGTGACCTTCTTAATCCTAATTAATATGGAATTAATTTGTGTTCTGTATTATGATTGCTTCTATCTGTTTACAAAGCAAGTTGAATGGTTGA 700  
CTTCACTGGAAGAAATAGGATTAATAATACCTTAATAAACAAACAAAGACATAATCTAACGAAAGATAGACAAATGTTGCTGTTCACTTAACCAACT

156 E V T F F N P N Y Y G I : C C F C I M I A F Y L F T T T K L N M L K 189

701 AAGTATTCTGTGATTGCAAGGCTTTGTTAATCTTTTGGTTGAACTTTTACTCAAAATCGAACTGCTTCTCTGCTATTATCGTGGAGCAATATCTA 800  
TTCAAGACACACTAACGTCGGAACCAATTAGAGAAACCAAACTTGAATGAGTTTACCTTGACGGAAGGACGATAATAGCGACCTCGTTAATAGAT

190 V F C V I A G F V N L F G L N P T O N R T A F P A : I A G A I I Y 222

801 TCTCTTACCACTATTAAAACTGCAAGGCTTTTGGCTTAGTATTGGGCTCTTCGGATTTGGTTGAGTTTCTCTTTCTAGTGATTTGGGAGTTTGA 900  
AGAGAAATGCTGATAATTTTGAACCTTCGGAAACCGAATCATAAACCCAGAAAGGCTAACCAAACTCAAGGAGAAAGATCACTAAACCTCAAGCT

223 L F T T I K N M K A F W L S I G V F A : G L S F L F S S D L G V R 255

901 ATGGGTACTTTAGACTCTCTATGGAAGAACGCAATTTCTATCTGGGATGCTGGGATGGCTTGTAAAGCAAAATCTTTTGGGCTGAAGGGCCATTGA 1000  
TACCCATGAAATCTGAGAAGATACCTTTTGGCTAAAGATAGACCTACGACCTACCGGAACAAATCTTTTAGGAAAAACCCACTTCCCGGTAAGT

256 M G T L D S S M E E R : S I M D A G N A L F K O N P F W G E G P L T 289

1001 CCTATATGCACTCTTATCCTCGGATACATGCTCTTATCATGAACATGCCACAGCTTTATATTGATACGATCTGAGTTACGGAATTTGGGTACCAT 1100  
GGATATACGTGAGAAATAGGAGCTATGTACGAGAAATAGTACTTGTACGGGTGTGAGAAATATAACTATGCTAAGACTCAATGCTTAAACCCATGGTA

290 Y M H S Y P R I H A P Y H E H A H S L Y I D T I L S Y G I V G T I 122

1101 TTTATTAGTTTGTCTTGTCTGCTCTGTTCTGTTGATGATGATATGAGTCAGGAGTCGGGAAACGTCCGATATCGGCTTTATCTATCTTTCTTT 1200  
AAATAATCAAAACAGAACACAGGAGCAAGCGAACTACTACCTATACTCAGTCTCAGCCCTTTGCAAGGCTAATAGCCGGAATAGATAGAAAGGAA

323 L L V L S S V A P V R L M M D M S O E S G K R P I I G L Y L S F L 355

1201 ACAGTGTTTGTGTGCAAGAAATTTTGAATTTGCTTTGCTGATTGAGTCAGGCTTTATTTTCTGCTAGTTATGTCAGTATTCATTTGGCTTTA 1299  
TGTCAACCAACGACAGTCTCTTAAAACTGAACCGAGAGAACCTAAGTCAGTCCGAAATAAAGAACGATCAATACAGCTCATAGGTAACCGAAAT

356 T V V A V M S : F D L A L F W I O S G F : F L L V M C S I P L A L 388

gsp222

Fig. 14

(SEQ ID NO: 41) 1 AAGGAGTGAACATCTGGCTCGGTACTTCAATTGATGAAAGTATGCGTGATGAAATTCGTGTAACAGTTGTCCCAACGGGTGTCGTCAAGACCGCGTAGA 100  
(SEQ ID NO: 42) TTCCTCACTTGTAGACCGAGCCATGAAGTTAACTACTTTCATACGCACTACTTTAAGCACATTGTCAACAGCGTTGCCACAAGCAGTTCTGGCGCATCT

101 AAAGCTTGTGGCTCCACAAGCTAGATCTGCTACTAAGTACCGTGAGACAGTGAACAGCTCATTCAATGGCTTTGATCGTCAATTTGATATGGCAGAA 200  
TTTCCAAACCGAGGTGTCGATCTAGACGATGATTGATGGCACTCTGTCACCTTTGCTCGAGTAAGTGTACCGAACTAGCAGTAAACTATACCGTCTT

201 ACAGTTGAATTGCCAAAAAATAATCCACGTCGTTTGGAAACCACTCAGGCATCTGCTTTGCTGATTGGGATCTTCCCGTGAATCGATTGTTCCGTACAA 300  
TGTCAACTTAAACGGTTTGTGTTAGGTGCAGCAACCTTGGTTGAGTCCGTAGACGAAACCACTAAACCTAGAGCGGCCTTAGCTAACAGCATGTT

301 CAGATTCACTCGTTCCTCAAGTCGAGCGCTTGAAGCCCAATTTCAAGATGAAGATGAATTGGATACACCTCCATTTTCAAAAAATCGTTAAGTAAA 400  
GTCTAAGTCAGCAAGAGGTGAGCTCGCGAACTTCGGGGTTAAAGTGTCTACTTCTACTTAACTATGTGGAGGTAAAAAGTTTTAGCAATTCAATT

(SEQ ID NO: 40) 1 M 1

401 TGAATGTAAAAGAAAAATACAGAACTTGTGTTTCCAGAAAGTTGCAGAGGCTAGTCTGAGTGCTCATCCAGACAGTGGTTCCGTCTCTGTCAATGCAAGTTA 500  
ACTTACATTTTCTTTATGCTCTTGAACAAAAAGCTCTTCAACGCTCTCCGATCAGACTCAGGATAGCTCTCTCACCAGCCAGAGACAGTAACTCAATTA

2 N V K E N T E L V F R E V A E A S L S A H R E S G S V S V I A V : 34

501 CAAGTATGTAGATGTACCGACAGCGGAAGCTTCTTCCGCTAGGTGTTCAATCATATCCGTGAAAAATCGGTAGATAAGTTTCTGAAAAAATATGAAGCT 600  
GTTCAATACATCTACATCGCTGTCGCTTCCGAAACGAAGGCGATCCCAAGTACTATAGCCACTTTTAGCACATCTATTCAAGACCTTTTATACCTCGA

35 K Y V D V P T A E A L L P L G V H H I G E N R V D K F L E K Y E A 67

601 TTAAGATCGAGATGTGACTTGGCATTGATTGGTACCTTGCAGAGAGCTAAGGTGAAGATGTCAATCAATACCTTGATTATTTCCATGCAATGGACT 700  
AATTTTCTAGCTCTACACTGAACCGTAACTAACCATGGAAAGCTTCTGCACTTCACTTCTACAGTAAGTTATGCAACTAATAAAGGTACGTAAACCTGA

68 L K D R D V T M H L : G T L Q R R K V K D V I O Y V D Y F H A L D S 101

701 CAGTAAAGCTAGCAGCGGAAAAATCAAAAAAGAAAGTACCCAGTCACTCAAGTGTCTTCTCAAGTAAATATTTCTAAGAGAGAAAGCAAAACCGGTTTTCT 800  
GTCAATTCGATCGTCCCTTTAAGTTTTTCTTCACTGGCTCAGTATTCAGAAAGGAAGTTCATTTATAAGATTTCTTCTTTCTGTTGTCGCAAAAG

100 V K L A G E : O K R S D R V I K C F L C V N : S K E E S K H G F S 134

801 GAGAGAGGAAGTGTGGAAATCTTCCAGAGTATGCCAGACTAGATAAGATTGAATATGTTGTTTAAATGAGATGGCAGCTTTGAGGCTAGCAGTGAG 900  
CTCTCTGTTGACGACCTTAGAACGGTCTCAATCGGTCTGATCTATTCTAACTTATACAAACCAATTAAGTCTACCGTGGAAAACTCCGATCGTCACTC

135 R E E L L E : L P E L A R L D K : E Y V G L M T M A P P E A S S E 167

901 CAGTTGAAAGAGATTTTCAAGCGCGCCCAAGATTTCAAAAGAGAAATTCAGAGAAACAAATTCAAAATATGCTTTAGAGCACACTGGCGCGCTTAC 999  
GTCAACTTTCTCTAAAAGTTCCGCGCGTTCTAAATGTTTTCTCTTAAAGTCTCTTCTTAAAGTTTATACCGAAATCTCGTGTGACCCCGCGCAATG

168 C L K E : F K A A C D L O R E : C E K C : P N M P L E N T G G R Y 200

gop228)

Fig. 15

(SEQ ID NO: 44) 1 GTACTCCAGTCCACTTTAGCAGTAAGTTATTATTACTTTAATCAGCCCAATTTCTTGTCTTGAATCAGATTTTGGTAGGTAGTTTGGTAATTC 100  
(SEQ ID NO: 45) CATGAGGGTCAGGTGAAAAATCGTCATTCAAAATAATAAATGAAAAATTAGTCGGTGTAAAGAAACAGAACTTAGTCTAAAACCATCCATCAAAACCATTAAGA  
(SEQ ID NO: 43) 1 T P S P L L A V S L L F T F N Q P Q F L V L N Q I L V G S L V I L 33

101 ACTTATTGCATATATAGTTGTAAAAATCCCATTTTCTTATAGAATGCTACGTGCTATTTTATTAGTCTTGATGATGAGATCGAAGATCCACCAAGAAAGT 200  
TGAATAACGTATATATCAACATTTTAGGGTAAAGAAATATCTTACCATGCAGGATAAAATAAATCACAACACTACTCTACCTTCTACGTCTTTCTCA  
34 L I A Y I V V K I P F S Y R M V R A I L F S V D D E M E D A A R S 66

201 ATGGGTGCTTACCTTTTATCTATGATGAAGGTATCATTCATTTATTTTACGGTGTCTCTCTGTTATTGCTTTAACTTTAACTTTTATTAA 300  
TACCCACCAAGTGGAAAAATATCATACTACTTCAATAGTAAGTAAATAAATGCCCCAACAGAGAGACAATAACGAAATTTGAAATTGAGAAATAAT  
67 M G A S P F Y T M M K V I I P F I L P V V L S V I A L N F N S L L T 100

301 CTGACTTCGACTTATCTGTATTCCTTACCATCCCTAGCTCAACCATTAAGTATTAAGATTGATCTGACAGGTGATGAACAGCAACATCTAATGCACA 400  
GACTGAAGCTGAATAGACATAAGGAAATGGTAGGGGATCGAGTTGGTAATCCATAATGCTAAGCTAGACGTCCACTACTTTGTCTGTAGATTACGTGT  
101 D F D L S V F L Y M P L A O P L G I T I R S A G D E T A T S N A Q 133

401 AGCTCTGGTATTTGTTTATACAAATGTTCTGATGATTATTTCTGGAACGGTATTACTTCAACAAGACCCGGGCGTAAAGTAAGCAATAATCATGA 500  
TCGAGACCATAAACAAATATGTTAAAGAGCTACTAATAAGACCTTGCCATAATATGAAGTGTGTTCTGGCCCCGATTTCACTCTTTATTAGTACT  
134 A L V F V Y T I V L M I I S G T V L Y F T Q R P G R K V R K \* 164

501 CAGCCACTAGTCTTGGGTATCAATATTGAATAGTTGTGAGGATTTGTTTATCAGTAGTCATTGGTAGTATAATGGTTTAGAGAGAGGGAGCAATC 600  
GTCCGTGATCAGAACCCCAATAGTTTATACTTTATCAACAGTCTTAACAAATAGTCATCAGTAACCATCATATTAAACCAATCTCTCTCCCTCGTTAG  
601 CCAGCCTGCAGGCATCCGAATTTATGATTTGTTGTTAGCTGCATGTTGATTATGATGACGAATGAATACGTATCTTATAAATTTGGACAGGAGAT 700  
GTCGGACGTCCGTAGGCTTGAATATCATAAACACAGATCGAGTCAAACTAATACTACTGCTTACTTATGATAGAAATTTAAACCCCTGCTCTTA  
701 CCTACACCATTAGGAGCTCAAGTTATATCAGGTGTGGGTTTTAGGCGGTGCAACGATCTTATTACAGATAAAAAGAAAAATACAGGTCTGACAACTG 800  
GGATGTGTAATCCTCGAGTTCAATATAGTCCACACCCAAAGATCCGGGACCTTGCTAAGAAATAATGCTATTTTTCTTTAATGTCCAGACTGTTGAC  
801 CAGCAGGCATTTGGGCTTCGGCAGGAATTGGATTAGCTATTGGAGTAGGTTTTATGAGGGAGCTCTTTAGTAGCCATTTCTGTTGGGGTGTGATATC 900  
GTCTGTCGTAAACCCGAAGCGTCTTTAACTAATCGATAACCTCATCCAAAAATACTCCCTCGAGAAAAATCATCGGTAAAGACAAACCCACACTATAG  
901 CATGTTCCAACTCAAAAAATATCTGCAAAATCGTTCTAAAAATGATTGAATTGTATATAGTAGTTAAATCCTTTAG 978  
GTACAGGTTGGTGATTTTTATAGACGTTTTAGCAAGATTTTACTAATTAACATATATCATCAATTTAGGAAATC

g9p273

Fig. 16

(SEQ ID NO: 47) 1 CAATGTGTTCCCGAACTTTTACAAAACTCTTCTGAAAAAGAGTTGGAACACTCAAGACCAATTTGGTCAAAATAGGATGGTTGTGGTTGATGATG 100  
(SEQ ID NO: 48) GTTACACAAGGGCTTGAAAACTTTTGTAGAGGACTTTTCTCAAGCTTGTGAGTTTCTGGTTAAACCACTTTATCCTACCAACCACTACTAC 100  
(SEQ ID NO: 46) 1 M M 2

101 CACAGGATTACACAAGAGTTGAAAAAGGTCGAGCTGTCTTCTACCTACAGAGCTGTTTATGGTCTTTTCCAGGCCTAGATGAAAAAGCAGTTG 200  
CTGTCTTAATCTGTTCTCAACCTTTTCCACCTCGACAGCAAGATGGATGTCTCTGACAAATACCAGAAAAAGGTTCCGGAATCTACTTTTCTGTCAC 200  
J D R I R Q E L E K G G A V V L P T E T V Y G L F S K A L D E K A V D 36

201 ACCATGTTTACCAACTCAAAAGTCTCTAGAGATAAGGCACTCAATCTCAATATCGCTCTTTCGAGGACATCTGCACCTTTCAAAGAAACAGCCAGC 300  
TGGTACAAATGGTTGAGTTTGCAGCAGGATCTCTATTCCGTGAGTTAGAGTTATAGCCGAGAAAGCTCCTGTAGAACGTGAAAAAGTTTCTTAGTCCGTCC 300  
37 H V Y Q L K R R P R D K A L N L N I A S P E D I L H F S K N Q P A 69

301 TTATCTACAAAACTTGTAGAGACCTTTTTCAGGTCCTCTGACCATTAATCTCGAAGCCAAATGACCGAGTTCCCTATTGGGTAAATTCGACCTTGCA 400  
AATAGATGTTTGTGAACATCTCTGAAAAACGGTCCAGGGAACTGGTAATAAGAGCTTCGGTTACTGGCTCAAGGGATAAACCAATTAAGACTGGAAAGCT 400  
70 Y L O K L V E T F L P G P L T I I L E A N D R V P Y W V N S D L A 102

401 ACTATTGGATTTCCGATGCCAGTCACTCTATCACTGGATTTAATTCGAGAGACAGGTCCTTGAATGGGCGCTCTGCCAAATCTCAGGTCAGGCCAA 500  
TGATAACCTAAAGCCTACCGGTCAGTGGGATAGTGTGACCTAAATTAAGCTCTCTGTCCAGGGAATTAACCCGGCAGACGGTTATAGAGTCCAGTCCGTT 500  
103 T I G F R M P S H P I T L D L I R E T G P L I G P S A N I S G Q A S 136

501 GTGGGTAAACCTTTGAACAAATTCGAAGGATTTTGACCAAGAGGTTCTGGGTCGGAAGACGATGCTTTCTAACTGGACAGGATCAACTATTGTGGA 600  
CACCACATTGGAAACTTGTTTAAGACTTCTTAAACTGGTTCTCAAGACCCAGACCTTCTGCTACGAAAAAGATTGACCTGTCTTAAGTTGATAACACCT 600  
137 G V T F E Q I L K D F D Q E V L C L E D D A F L T G Q D S T I V D 169

601 TTTGTCTGGAGACAAGGTTGAAAACTTACCCAAAGCCGCAATTAACGAGAGATATTCTTGTCTGGTGGCCAGAGATTTCTTTTGAGGAGGCTTGAATG 700  
AAAACAGACTCTGTTCCACTTTTGAATGGGTTCCGGTTAATTTGGTCTCTATAAGAACGAGCCAAACGGTCTCTAAGAAAACTCTCCGAACTTTAC 700  
170 L S G D K V K I L P K A Q L N E K I F L L G C Q R F L L R R L E M 202

701 CTAAGAGATTGCAAGAAACAGATGTGAAGCGATATGTGACATCAACCAAGAGGCTTTGGGTTATACTTTAGTCCAGAGGAAACGGCTAGCCAACTAG 800  
GATTTCTTAAAGCTTTCTTTGTCTACACTTTCGTATACACTGTAGTTGGTCTCCGAAACCAATATGAAAAATCAGGTCTCCTTTGCCGATCGGTTGATC 800  
203 L R D L Q E T D V K A : C D I N Q E A L G Y T F S P E E T A S O L A 236

801 CTAGACTGTCTCAGGATTCCTCATTTCTCTACTTGGCTATGAGGATGCAGCTAATCATGTCTTACTTGGATATGTCCACGCTGAAGTTTACGAATCACT 900  
CATCTGACAGCTCTAAGGGTAGTAAAGGATGAACGATACTCTACGTGATAGTACAGAAATGAACCTATACAGGTGGGACTTCAATGCTTACTGA 900  
237 R L S Q D S H H F L L G Y E D A A N H V L L G Y V H A E V Y E S L 269

901 CTATTCAAAGCAGGATTTAATATCTTAGCTTTAGCAGTTTCACTCAAGCCCAAGGTCAAGGTATCGTAAAGCTTTACTACAAGGTTGGAAACAGAA 1000  
GATAAGGTTTCTCTTAAATATAGAAATCGAAATCTCAAGTGGAGTTCCGGTCCAGTTCCATAGCCATTTTCAATGATGTTCCCAACCTTTGTTCTT 1000  
270 Y S K A G F N I L A L A V S P O A Q G Q G I G K S L L O G L E Q E 302

1001 GCCAAAGATGTGGTTATGGGTTTATCCGTTAAATTTCTGCAATCATGTTGGGTCATGCTTTTATGAAAAAGTTGGCTATATCTTGTGATAAA 1100  
CGGTTTTCTACACCAATACCAAAATAGCGGAATTAAGACGGTTAGTAGCAGACCCAGGATACGTAATACTTTTCAACCGATATGAACACTATTTT 1100  
303 A K R C G Y G F : R L N S A N H R L G A N A F Y E K V G Y T C D K M 336

1101 TGCAGAAACGGTTTATTCGCACTTTTACTTTCATTTCTTATGTAATAATCAAACTAATGGACTAGTCAACAATAAAGGAGAAGACCTATGATTTTGG 1200  
ACGTCTTTGCCAAATAAGCGTAGAAAAATCAAACTAAAGAAATAACATTTTACTTTGATTACCTGATCAGTGTGTTATTTCTCTTCTGGATACTAAAAAC 1200  
337 C K R F : R I F \* 345



9ep286

Fig. 17 (Sheet 1 of 2)

(SEQ ID NO: 50) : AAGATAATAGAAAAAGAAATGTAACGAATGAGAGAAAAATGCCATTTGGAGATAATCGAAATCGTAAAAAACTATGTTGAGAAAAATACCTTGTTAT 100  
(SEQ ID NO: 51) : TTCTATTATCTTTATCTTACATTGCTTACTCTCTTTTACCGTAAACCTCTATTACCTTAGCATTTTTCGATACAACTCTTTTATTGGAAACAAATA

101 CGTGATTATCATGCTAGTAGCAAGTTTATTGGGAATTTTTCGAACTGCAATTTGGTGCTTCACTAATCTATAAAATGATTCAAGAAAAATTTAGTCACTG 200  
GCACTAATAGTAGCATCATGTTCAAATAACCTTAAAAACGTTGACGTTAACCAAGGAGTCATTAGATATTTTAACTAGTCTTTTAAATCACTGAC

201 GGATTTCCAGCCCTTTTAAAGTGAGAGAAAAATAGTATGTTTATAGATACAGCTAAGATTAAGGTCAGGCTGGTAAATGCTGGCGATGGTATGG 300  
CCTAAAGGGTCGGGAAAAATTTCACTCTTCTTTATTACTCATACAAAAATCTATGTCGATTCTAATTCCAGTTCCGACCATTAACACCGCTACCATACC

(SEQ ID NO: 49) : M F L D T A K I K V K A G N G G D G M V 20

301 TTCCCTTTCTGCTGAAAAATATGTCCTAATGGAGGCCCTTGGGGTGGTGGTGGTGGTGGAGGCAATGTCCTTCTGTTAGACGAAGGACTACG 400  
AACGGAAGCAGCACTTTTATACAGGGAATTAACCTCGGGAACCCCACTACCAACGAGCACTCCGTTACACGAGAAACATCTGCTTCTGATGC

21 A F R R E K Y V P N G G P W G G D G G R G G H V V F V V D E G L R 53

401 TACCTTGATGGATTTCCGCTACAATCGTCAATTCAGGCTGATTCCTGGTAAAAAGGGATGACCAAGGGATGCAATGCTGCTGCTGAGCACTTACA 500  
ATGGAATACCTAAAGGCGATTTAGCAGTAAAGTTCCGACTAAGACCACTTTTCCCTACTGTTTCCCTACGTACGAGCACCAGCACTCTGGAATCT

54 T L M D P R Y N R H F K A D S G E K G M T K G M H G R G A E D L R 86

501 GTTCGAGTACCAAGGTACGACTGTTCTGATGCGGAGACTGGCAAGGTTTAAACAGATTTGATTGAACATGGGCAAGAAATTTATGTTGCCCAAGGTC 600  
CAAGCTCATGGTCTCATGCTGACAGCACTACGCTCTGACCGTTCCAAAAATTTGCTAAACTAACTGTACCCGTTCTTAAATAGCAAGGGTGCAC

87 V R V P Q G T T V R D A E T G K V L T D L I E H G Q E F I V A N G G 120

601 GTCTGCTGAGCACTGGAATATTCGTTTCCGACACCAAAAAATCTGCAACCGGAAATCTCTGAAAAAGCAACCAAGCTCAGGAACGTGAGTTACAAT 700  
CAGCAACCTGCACTTTATAGCAAGGCGCTGCTGTTTATAGGACCTGCGCTTAGAGACTTTTACCTCTGCTCAGTCTCTGCACTCAATGTTAA

121 R G G R G N I R F A T P K N P A P E I S E N G E P G Q E R E L O L 153

701 GGAATCAAAATCTTGGCAGATGTCGTTTAGTAGGATTCCCATCTGTAGGGAAGTCAACACTTTAAGTGTATTACCTCAGCTAAGCTCAAAATTTGGT 800  
CCTTGATTTTGAACCGCTACAGCCAAATCATCTAAGGCTAGACATCCCTTCAGTTGTGAAAAATCACAAATAAGGAGTCGATTTTAAACA

154 E L K I L A D V G L V G F P S V G K S T L L S V I T S A K P K I G 186

801 GCCTACCACTTTACCACTATTGTACCAAAATTTAGGTATGTTTCCGACCAATCAGGTGAATCTTTGCACTAGCCGACTTCCAGCTTTGATTGAAGGG 900  
CGGATGGTGAATGGTGAATACATGTTTAAATCCATACCAAGCGTGGGTTAGTCCACTTAGGAAACGTCATCGGCTGAACGGTCCAAACTAATCTCC

187 A Y H F T T : V P N L G M V R T O S G E S F A V A D L P G L I E G A 220

901 CTAGTCAAGGTGTTGTTGGAACTCAGTCTCTCGTCACTCGAGCGTACAGTGTATCTTCACTCATTCATATGTCAGTACGCAAGGCGGTGA 1000  
GATCAGTTCCACAAACCAACCTTGAGTCAAGGAGGAGTGTAGTCCGATGTGCAAAATAGGAAGTGTAGTAATATACGTGATCGCTTCCGCACT

221 S C G V G L G T C F L R H I E R T R V : L H I I D M S A S E G R D 253

1001 TCCATATGAGGATTACCTAGCTATCAATAAGAGCTGGAGTCTTACAATCTTCCGCTCATGGAGCGTCCACAGATTATTGTAACTAATAAGATGGACATG 1100  
AGGTATACTCTAATGGATCGATGTTATTTCTGACCTCAGAAATGTTAAGCGGAGTACCTCGAGGTGTCTAATAACATTGATTATTCTACCTGTAC

254 P Y E D Y L A I N K E L E S Y N L R L M E R P Q : I V T N K M D M 286

1101 CCTGAGAGTACGAAAAATCTTGAAGAAATTAAGAAAAATTTGGCTGAAAAATTAATGATGAATTTGAAGAGTTACAGCTATCTTCCCAATTTCTGGATTGA 1200  
GGAATCTCAGTCTTTTGAACCTCTTAAATCTTTTAAACCGACTTTTAACTACTTAACTCTCAATGGTGGATAGAAGGGTTAAAGACCTAAT

287 P E S O E N L E E F K K Y L A E N Y D E F E E L P A I F P I S G L T 320

1201 CCAAGCAAGCTTGGCAACACTTTAGATGCTACAGTGAATTTGTAGACAAGACCAAGAAATTTGCTCTACGAGGAGTCCGATATGGAAGAAGAAGT 1300  
GTTCTGCTCCAGACCTTTGTGAAAAATCTACGATGTCAGCTTAACAATCTGTTCTGTGCTTAAAAACGAGATGCTGCTCAGGCTATACCTTCTTCTCA

321 F O C L A T L L D A T A E L L D K T P E F L L Y D E S D M E E E V 353

Fig. 17 (Sheet 2 of 2)

1301 TTACTATGGATTGACGGAAGAAAAAGCCCTTTGAAATTAGTCGTGATGACGATCGGACATCGGTACTTTCTCGTGAAAAAATCATGAACTCTTTAAAT 1400  
AATGATACCTAAACTGCTTCTTCTTTTCCGAAACTTTAATCAGCACTACTGCTACGCTGTACCCATGAAGACCACCTTTTTCAGTACTTTGAGAAATTA

354 Y Y G F D E E E K A F E I S R D D D A T M V L S G E K L M K L F N 386

1401 ATGACCAACTTTGATCGTGATGAATCTGTCAATGAACTTTA 1441  
TACTCGTTGAACTAGCACTACTTAGACAGTACTTTGAAAT

387 M T N F D R D E S V M K L 399

gcp311

Fig. 18

(SEQ ID NO: 53) 1 TCGAATGCCCTTAAGAAAACAAATGAAAATCAAGAAAACAGTANGACAAGTTCTTTGCTCTTATGAATTATTAGAAATGAAGAAAGAAAGCATATTAT 100  
 (SEQ ID NO: 54) ACCCTACGGGAATCTTTTGTAACTTTAGTTCTTTTGTCTGTTCAAGAAAACAGAACTACTTAATAATCTTTACTTCTTTCTTCTATAATA 100  
 (SEQ ID NO: 52) 1 M 1

101 GGCTGAAGAAAGAGTAGAACCAAAACCAATTGACCTTGGTGAATATAAAATTTGGTTTCCATGACGATGTAGAGCCTCTCTTATCGACAGGAAAAGGACTC 200  
 CCGACTTCTTTCTCATCTTGGTTTGGTTAACTGGAACCACTTATATTTAAACCAAGGTACTGTACATCTCGGACAGAAATAGCTGTCTTTTCTGAG

2 A E E R V E P K P I D L G E Y K F G F R D D V E P V L S T G K G L 34

201 AACGAAGGTGTTATTCTGGAATTATCTGCTAAGGGTGAGCCTGAGTGGATGTTGGAGTTCCGTTTGAAATCTTATGAAACCTTCAAAAAATGCCCA 300  
 TTGCTTCCACAATAAGCACTTAATAGAGCAGATTCCCACTCGGACTCACCTACAACTCAAGGCAAACTTCAGAATACTTTGGAAAGTTTCTTACGGGT

35 N E G V I R E L S A A K C E P E W H L E P R L K S Y E T P K K M P H 68

301 TCGAACTCTGGGAGCAGACTTGTGAGAGTTGACTTTGATGACTTAATCTACTACCAAAAACCACTGCAAAACCAAGCCGCTTCTGGGATGATGTACC 400  
 AGCTTTGAACCCCTCGTCTGAACAGTCTCTAACTGAACTACTGAATTAGATGATGGTTTTTGGTAGACTGTTTGGTGGGCAAGAACCTTACTACATGG

69 Q T W G A D L S E I D F D D L I Y Y O K P S D K P A R S W D D V P 101

401 TGAAGAGATTAAAGAAACCTTTGAACGTATCGGGATTCCGAAGCTGAACGTGCTTATTTAGCAGGGGCTTCTGCCAGTACGAGTCAGAAGTGGTTTAC 500  
 ACTTTTCTAATTTCTTTGGAACCTTGCATAGCCCTAAGGCTTTCGACTTGCACCAATAAATCGTCCCGAAGACGGGTCACTGCTCAGTCTTCCACCAATG

102 E K I K E T F E R I G I P E A E R A Y L A G A S A O Y E S E V V Y 134

501 CACAACTCAAGCAAGCTTCCAAAATAGCTATTATCTTTACAGATACAGATTCCCACTCAAGCAATACCCAGACTTATTTAAACAATACTTTGCCA 600  
 GTGTTGTACTTCTTCTCAAGGTTTTTAATCCATAATAGAAATGTCTATGCTCAAGGCTGAGTTCTTATGGGTCTGAATAAATTTGTTATGAAAGCT

135 H N M K E E F O K L G I I F T D T D S A L K E Y P D L F K O Y F A K 168

601 AGTGGTACGCGGACAGATAACAAGTTGCGAGCCTCAACTCAGCAGTATGGTGGGTGGAACTTTTATCTACGTGCCAAAAGGTGTCAAGGTAGATAT 700  
 TCAACCATGGCGGCTGTCTATTGTTCAACCGTGGGAGTTGAGTGTCTATACAGCCCACTTGAAAATAGATGCAGGTTTTCCAGAGTTCATCTATA

169 L V P P T D N K L A A L N S A V W S G G T F I Y V P K G V K V D I 201

701 TCCACTTCAAACTTATTTCCGTATCAATAACGAAAATATAGGTGAGTTGCAAGCTACCTTGATTATGTTGATGAGGGAGCAAGCTCCACTACGTAGAA 800  
 AGGTGAAGTTTGAATAAAGGCATAGTTATTTGCTTTTATATCCAGTCAAGCTTGCAATGGAATAGCAACTACTCCCTCGTTCGAGGTGATGATCTTT

202 P L Q T Y F R I N N E N I G O F E R T L I I V D E G A S V H Y V E 234

801 GCATGTACAGCACCAACATATTCAAGCAATAGCTTACAGCTGCCATTTGTAGAAATTTTGGCTTTGGACGGAGCTTATATGGTTATACAACTATCCAAA 900  
 CTTACATGTCTGGTTGTATAAGTTCTTATCGAATGTGCAAGGTAACTCTTTAAAAACGAAACCTGCTCGAATATACGCAATATGTTGATAGGTTT

235 G C T A P T Y S S N S L H A A I V E I F A L D G A Y M R Y T T I O N 268

901 ACTGGTCTGATAACGTCTATAACTTGGTAACAAAGCTGCTAAGGCTCAAAAGGATGCCACTGTTGAGTGGATTGATGGAACTTGGGTGCCAAAACGAC 1000  
 TGACCAGACTATTGAGATATTGAACCAATTGTTTCCAGGATTCCGAGTTTCTACGGTGACAACTCACCTAACTACTTGAAGCCACGGTTTTGCTG

269 W S D N V Y N L V T K R A K A O K D A T V E W I D G N L G A K T T 101

1001 TATGAAATATCCATCTGTTTACCTTGTATGGAGCAAGGAGCGGTGGTACCATGCTCTCTATCGCCTTTGCTAATGCAGGGCAACACCAAGACACGGGTGCT 1100  
 ATACTTTATAGGTAGACAAATGGAACCTACTCTTCTCGCCACCATGGTACGAGATAGCGGAAACGATTACGTCCCTTGTGGTTCTGTGCCCCACGA

302 M K Y P S V Y L D G E G A R G T H L S I A F A N A G O H O D T G A 334

1101 AAGATGATTACAAATGCTCCACATACCAGCTCGTCTATTGTCTAAATCCATCGCTAAAGGTGGAGGAAAGGTTGACTACCGTGGCAAGTCACCTTTA 1200  
 TTTACTAAGTGTACGAGGTGTATGCTCGAGCAGATAACACAGATTTAGGTAGCGATTTCCACTCTTTCCAACTGATGGCAGCTGTTCACTGGAAAT

335 K M : H N A P H T S S S : V S K S : A K G G G K V D Y R G O V T F N 368

1201 ACAAGAACTCTAAGAAATCTGTTTCCACATTTGAATGTGATACCATTTATCATGGATGACCTTT 1263  
 TGTCTTTGAGATCTTTAGACAAAGGGTGAACCTTACACTATGTTAATGTAACCTACTCGAAA

369 K M S K K S V S H : E C D T I I M D D L 388

gsp2262 Fig. 19

(SEQ ID NO: 56) 1 AGCTGGAAATTTATGAGCAAGTATCCTATCTTAAGAAGCAAGAGTGTATCTAACTCGTTATAATGAAGTCAAACTGAAAACAGCAACTTTAATCTTA 100  
(SEQ ID NO: 57) TGGACCTTAAATACTCTTTCATAGGATAGAAATTTCTTCTTTCACAAATAGATTGAGCAATATTACTTCAAGTTTGACTTTGTCTTGAATTAGAAT  
(SEQ ID NO: 55) 1 A G I Y E Q V S Y L K E G R S V Y L T R Y N E V O T E T A T L I L 33

101 GGAGCTATTGTGGGATAGCTACTTCCTTCTTACTCTTTTATTCTGTCAATCTTCTATATTTCGAGCAATTCGCGGAGATATCTTGATTAAACGAATTT 200  
CCTCGATAACACCCCTATCGATCAAGGAACAAATGAGAAATAGACAGTTAGAAGATATAAAGCTCGTTAAGGCGGCTCTATAGAACTAAATTTGCTTAAA

14 G A I V G I A S S L L L F Y S V N L L Y P E O F R R D I L I X R I S 67

201 CAGGTTTACGATTTTTTGAACACATGCTCAGTATATGGTTAGTCAATTTGCCAGTTTGTATTTGGTGTAGTCTCTTTATTTAAGCAGTCGAGACTT 300  
GTCCAAATGCTAAAAAATTTGTGTACGAGTCATATACCAATCAGTTAAACGGTCAAAACATAAACCAAGATCAGAGAAATAAAATTCGTCAGCTCTGAA

68 G L R F F E T H A Q Y M V S O F A S F V F G A S L F I L S S R D L 100

101 GGTGATTGGCTTGCTCACTTTATTAGTCTTTCTAGCTAGTGCAGTTTGAAGGCTTTACCGTCAAGCGCAGAAAGAAATCTCGTCTTTCTATGACAATTATG 400  
CCACTAACCGAACCGAGTGAAATAATCAGAAAGATCGATCAGTCAAAACTGCGAAATGGCAGTTCCGCTCTTTCTTAGAGCACAAGATACTGTTAATAC

101 V I G L L T L L V F L A S A V L T L Y R Q A Q K E S R V S M T I M 133

401 AAAGGAAAATAGGATGATTGAACTAAAGAAATATCTAAAAAATTTGGAAGCCGTCAGCTATTTTCAGATACGAATCTTTA 481  
TTTCTTTTATCCTACTAAGTTCATTTCTTATATAGATTTTAAACCTTCGGCAGTCGATAAAAGTCTATGCTTAGAAAT

134 K G K • 137

gcp1187

Fig. 20

(SEQ ID NO: 59) 1 TTTTATCTAGTACAGTATATTTATTGCGGTGTCGCAATATTCAATCCATCCAAATGTATTAGAAATGGATCTTAGTTTACTTCAAGATATGACGACTGG 100  
AAAATAGATCATGTCTATATAAATAACGGGACAGCGGTTATAAGTTAGGTAGGTTTACATAATCTTACCTAGAATCAAAATGAAGTTCTATCTGCTGACC  
(SEQ ID NO: 60) M T T G 4  
(SEQ ID NO: 58) 1

101 AGTATATTGCTTTCCGTTCAATATATATGTTCTTTTATTGATGAATAACTATTTAAATAGGTTGGAGTGTGCGATTGCTGAAATCAATTAAG 200  
TCATATAACGAAAGCCAGTGTATATATAACAGAAAAATAAACTACTTATTGATAAAATATCCAACTCACAGCGTAAGCAGACTTTAGTTAATTC  
5 V Y C P P F T Y I L F F F Y L N W N Y F N R L E C R I R L K S I K 37

201 CACTTTACCAAGTTTACTTTCAAATTAGCAGCTCTTAGTAGCGGGATTGGACGGCGACTTTATTTTATTGATTTTCTAATTGCATTTAGTAATGGTT 300  
GTGAAATGGTCAAAATCAAGTTTAATCGTCGAGAAATCATGCCCTAAACCTGCGCTGAAATAAAAATAACTAAAAAGATTAAAGTAAATCATTACCAA  
38 H F T S F S F K L A A L S T G I W T A T L F L L I F L I A F S N G F 71

301 TTAGCTTCTCTTTGGAGATAAAGCAAGTTGATTTTAAAGAGAAATTTATCGTATAAGTATTGCAAAATGCTAGTTTCTTATAGGATTTTCTTCTC 400  
AATCGAAGAGAAACCTCTATTTCTCCTCAACTAAAAATTTCTTTAAATACCATATTGATAACGTTTGTAGCATCAAGAAATATCTTAAAAAAGAG  
72 S F S L E I K E V D F L R E F Y G I S I A N N A S F F I G F F F S 104

401 TTATATAGCATACTATTTCTTTTATCCTTACTTACTATTAGCAGTTTCTTGGTTAAAAAATCAACATGAGCTTAGTATTCTGTTTACTTTTAA 500  
AATATATCGTATGATAAAGAAAAATAGGAATGAATGATAATCGTCAAAAGAACCAATTTTATTAGTTTGTACTCGAATCATAAAGACAAATGAAAAAT  
105 Y I A Y Y F F L S L L T I S S F S W F K K S N M S L V F L F T F L 137

501 TTTGTAGAATCTTATTCTGGATTATCAGTTGGACAAATGGGATAATGGATTATTGCCAATTTTTCAGTATATGCTAAATCCAAATCCGTATGCAATTGA 600  
AAACATCTTAGGAATAAGACCTAAATAGTCAACCTGTACCTTATTAACCTAATAACGGTTAAAAAGTCATATACCATTTAAGGTTAGGCATACGTAAT  
138 F V E S L F W I Y Q L D N G I I G L L P I F O Y N V N S N P Y A L I 171

601 TTTATTGCTTACATTACTATCTATCATAAATCCATTGACTGTATTTCTGTTTATAGAAATCGGAGGAGAGTGTAAAAAGTTGAAAAATGGGAAAGTTAAG 700  
AAATAACCGAATGTAATGATAGATAGTATTAAGGTAACGACATAAAGACAAGTATCTTGACCTCTCTCACATTTTCAACCTTTACCTTTCAATTC  
172 Y W L T L L S I I I P L T V F S V H R N W R R V \* 196

gop47

Fig. 21 (Sheet 1 of 2)

(SEQ ID NO: 62) 1 AGGGAACAGAAAAATTTCAAGTTTCTGTATATAATAGAACTCTGTATATAAGGAGGTAAATCATOGAATTAGTGCATGGAAATTTCAAACATTTTATCC 100  
(SEQ ID NO: 63) TCCCTTGTCTTTTAAAGTCCAAAGCACTATATTATCTTCAGACATATATTCCTCAATTTAGTACCTCAATCAGTACCTTAAAGTTGTGTAAGATAGG  
(SEQ ID NO: 61) 1 M E L V H G I S T H P I Q 11

101 AATCAAAAAAGTTTAAACAAACAAATACCGTGCGTTTACCCTCCATTATCCCTGATACGATTGCAGGTACATGTTGAGTGCAGTATGCTAGA 200  
TTAGTTTCTTCAAATTTGTGTTTAAATGGCAGCAAAATGGCGAGGTAATAGGGAATCTATGCTAACGTCAGTGTACAACTCAGTTTCATAGCATCT  
14 S X K F K T N K I T V R F T A P L S L D T I A G H M L S A S M L E 46

201 GACTGCTAATCAGATGTACCCCACTTCTCAAGATTTGAGGAGACACTTGGCCAGTCTATACGGTACAGATATGTCACCAATTTGTTTCAAGAGAGGGCAA 300  
CTGACGATTAGTCTACATGGGGTGAAGAGTTCTAAACTCTCTGTGAACCGGTGAGATATGCCATGTCTATACAGTTGGTTAAACAAAGTCTTCTCCCGTT  
47 T A N Q M Y P T S O D L R R H L A S L Y G T D M S T N C F R R G Q 79

301 AGCCACATTTATAGAATTGCACTTACCTATGTTCTGTATGAGTTTAAAGTAGGAAAAAGGTGCTAACCTCTCAGATTTCGAACTCTTAAAAGAAACTC 400  
TCGGTGTAAATCTTAACTGTAAATGGATACAGCACTACTCAAAAATTCATCCTTTTGCAGATTGGAGAGTCTAAAACCTTGAACATTTCTTTGAG  
80 S H I I E L T F T Y V R D E F L S R K N V L T S O I L E L V K E T L 113

401 TTTTTTCACTCCAGTGTGATAATGGGTTTGTATCCGGCTTATTTGAAATGAGAAAAACAAATGCTAGCAAGTTTAGCAGCTGATATGGATGATTC 500  
AAAAAAGTGGGCTCATCACTATTACCAAACTAGGCGGAATAAATCTTAACTCTTTTGTAAAGATGTTTCAAACTCGACTATACCTACTAAG  
114 F S P A V V D N G F D P A L F E I E R K O L L A S L A A D M D D S 146

501 TTTTATTTTGCACATAAAGAAATGGATAAATGTTTTTTCATGATGAAGCTCTTCAATGGAATATAGTGATTTAGCAAACTCGTATTTTAGCTGAAACT 600  
AAAAATAAAGCTGTATTTCTTAACTATTAAACAAAAAGTACTACTGTCAGAAAGTTAACCTTATATCACTAAATGCTTTAGCATAAATCGACTTTGA  
147 F Y F A K K E L D K L F F H D E R L O L E Y S D L R N R I L A E T 179

601 CCACAAAGTTCTTATCTTGTCTCAAGAAATTTAGCCAAATGATCGAATAGATTTCTTTCTAGGTGATTTTAAAGAGGTGAAATTCAAAATGTAT 700  
GGTGTCTCAAGAAATAGAACAAAGGTTCTTAAAAATCGGTACTAGCTTATCTAAAGAAAAAGGATCCACTAAATTAATCTCAACTTTAAGTTTACATA  
180 P O S S Y S C F O E F L A N D R I D F F F L G D F N E V E I Q N V L 213

701 TAGAATCAATTCGCTTTAAAGGTCGAAAAGGAGATGTGAAGGTTCAATTTGTCAACCTTATCTAATATCTTTCAGGAAGGTATGGTTTCGGAATAATGT 800  
ATCTTAGTAAACCGAAATTTCCAGCTTTCTCTACACTTCCAAAGTCAACAGTTGGAATAGATTTAGGAAGTCTTCCATACCAAGCTTTTATCA  
214 E S F G F K G R K G D V K V Q Y C C P Y S N I L O E G H V R K N V 246

801 CGACAAATCCATTTTGAATTAGGTTATCATTACCGTTCTAAATATGCTGATGAGCAACATTTACCCATGATGTAATGAATGGTTTACTTGGTGGATTT 900  
CCCTGTTAGGTAAACCTTAATCCAATAGTAATGGCAAGATTATACCACTACTCGTTGTAAATGGGTACTAACATTAATTAACCAATGAACCACTAAA  
247 C O S I L E L G Y H Y R S K Y G D E O H L P M I V M N G L L G G F 279

901 GCTCACTCTAAGCTCTTTACAAATGTCCGTGAAAAATGCTGGATTAGCTTATACCAATTTCAAGTGAGCTTGATTTATTTAGTGATTTCTGAGGATGTATG 1000  
CGAGTGAGATTGAGAAATGTATACAGGCACCTTTACGACCTAATCGAATATGGTAAAGTCACTCGAACTAAATAAATCACTAAGAACTCCTACATAC  
280 A H S K L F T N V R E N A G L A Y T I S S E L D L F S G F L R M Y A 313

1001 CTGGTATCAATCGAGAAATCGTAACCAAGCTCGTAAATGATGAATAATCAACTGCTTGATTTAAAAAAGGTTATTTTACAGAGTTTGAAGTTAAATCA 1100  
GACCATAGTTAGCTCTTTAGCATTTGGTCCGAGCACTTTACTACTTATTAGTTGACGAATAAATTTTCCAAATAAATGTCTCAAACTCAATTTAGT  
314 C I N R E N R N O A R K M M N N O L L D L K K G Y F T E F E L N O 346

1101 GACCAAGCAATGATTCGTTGGTGGTTGTTACTTTCTCAAGATAATCAATCTTCATGATTGAAAGTCTTATCAAAATGCTTATTTGCAAAATCTTCA 1200  
CTGGTCTCTTTACTAAGCAACCAAGCAATGAAGAGTTCTATTAGTATAGAACTAACTTGCACGAATAGTTTACGCAATAAATCTTTTGAAGT  
347 T K E M I R W S L L L S O D N O S S L I E R A Y O N A L F G K S S 379

1201 GCAGACTTTAAAGTTGGATTGCAAGCTTGAAACAAATGCAAAAGATGCTATTTGTAGAGTACCTAATAATGTGAAACTCAAGCGATTTACTTTATGG 1300  
CGTCTGAAATTTCAACCTAAAGTTTCAAGCTTGTCTAACTGTTCTACGATAAATCATCTCATGATTATTACACTTTGATGTTGGTAAATGAATATCC  
380 A D F K S W I A K L E Q : D K D A : C R V A N N V K L O A I Y F M E 413

## Fig. 21 (Sheet 2 of 2)

1301 AAGGAATAGAATGACAAAGGTTGT:TTTGAACAAAAATACTATCCAGCTGTAAAAGAAAAGGTTTATCGAACTCGTTTGGCCAAACGATTGACAGTTGCT 1400  
TTCCTTATCTTACTGTTTCCAAACAAAACCTTCT:TTTATGATAGGTGACATTTTCTTTTCCAAATAGCTTGAGCAAAACCGTTGCTAACTGTCAACGA  
414 G I E • 417

gsp61

Fig. 22

(SEQ ID NO: 65) 1 GTTTTGTGACCATTTCAAAAGTCCTTAGCAGAGGAGGCTCTATCTTGAAGAGAAATTTATTACCTTTCAAAATCTGACTTTGGTATTTATTT 100  
(SEQ ID NO: 66) CAAAAAATCGGTAAAGTTTTCAGCAATCGTCTTTTCTTTCAGCAGATATGAAGCTTTCTTAAATAAATGGAAGTGTAGACTGAAACCATAAATAAA

101 TAGAGAAAAATTAAGTTCTCCCATGGTTTATCGAGAGGTTCTCTTTATCGCAATGAAGATTTAGTAGTGGAAATCTCGGAAATGACTCCCAAAACAAGT 200  
ATCTCTTTTAAATCAAGAGGGTACCAAAATACCTTCCAGGACAAATAGCGTTACTTCTAAATCATCACCTTAGACCTTTAACTGAGGGTTTGTTCAT

(SEQ ID NO: 64) 1 M V Y G E V P V Y A N E D L V V E S G K L T P K T S 26

201 TTTCAAAATAACCGAGTGGCGCTTAAATAAACAAGGAATTCAGTATTTAAGCTATCAATCATCAATTTATAGCTCGGACAAACGATTTTATATGATC 300  
AAAGTTTATGGCTCACCGCGAATTTATTTGTTCTTAAGGTATAAATCGATAGTTTAGTAGTTAAATATCGACGCTTGTTCGTAATAATATAGTAG

27 F Q I T E W R L N K O G I P V F K L S N H O F I A A D K R F L Y D Q 60

301 AATCAGAGGTAACTCCAAATAAAAAAGTATGGTTAGAAATCTGACTTTAAACTGTACAAATAGTCTTATGATTTAAAGAAAGTAAATCATCTTATC 400  
TAGTCTCCATTGAGGTGTTTATTTTTCATACCAATCTTAGACTGAAATTTGACATGTTATCAGGAATAGTAAATTTTCTTCACTTTAGTAGGAATAG

61 S E V T P T I K K V W L E S D F K L Y N S P Y D L K E V K S S L S 93

401 AGCTTATTGCAAGTATCAATCGACAGAGCATGTTTGTAGAAGGAAGAAATTTCTACATATTGATCAGGCTGGATGGGTAGCTAAAGAAATCAACTTCT 500  
TCGAATAAGCGTTTATAGTTAGCTGTTTCTGTACAAACATCTTCTCTTCTTAAAGATGTATAACTAGTCCGACCTACCCATCGATTTCTAGTTGAAGA

94 A Y S O V S I D K T M F V E G R E F L H I D O A G W V A X E S T S 126

501 GAAGAAGATAATCGATGAGTAAAGTTCAAGAAATGTTATCTGAAAAATATCAGAAAGATTTCTTCTCTATTATGTTAAGCAACTGACTACTGAAAAAG 600  
CTTCTCTATTAGCTTACTCATTTCAAGTTCTTCAATAGACTTTTATAGTCTTTCTAAGAAAGAGATAAATAACAATTCGTTGACTGATGACCTTTTC

127 E E D N R M S K V Q E M L S E K Y O X D S F S I Y V K O L T T G K E 160

601 AAGCTGGTATCAATCAAGATGAAAGATGATGCGAGGAGGTTTTCGAACTCTCTTATCTTATTATAGCGAAGAAAAATAAATGAGGGTCTTTATCA 700  
TTGACCATAGTTAGTTCTACTTTCTACATACGTCGGTCCGCAAACTTTGAGAGAAATAGAGATAAATAGCGTTCTTTTATTATTACTCCAGAAATAGT

161 A G I N O D E X M Y A A S V L K L S Y L Y Y T O E K I N E G L Y O 193

701 GTTAGATAGGACTGTAAATATGATTTGAGTCAATGATTTTCAGGTTCTTATAAACAGAGGGAAGTGGTAGTCTTCTAAAAAAGAGATAATAAA 800  
CAATCTATGCTGACATTTTATGATAGAGGTGAGTTACTAAAGGTCCAGAAATATTTGGTCTCCCTTACCATCAGAAAGATTTTCTTCTATTATTTT

194 L D T T V X Y V S A V N D F P G S Y K P E G S G S L P K K E D N K 226

801 GAATATTTTAAAGSATTAAATACGAAGTATCAAAAGAAATCTGATAATGTAGTCTAATCTATTGGGATATTACATTTCAACCAATCTGATGCCA 900  
CTTATAAGAAAATTTCTAAATTAATGCTTCTATAGTTTCTTAGACTATTACATCGAGTATTAGATAACCTATAATGTAAGTTTGGTTAGACTACGGT

227 E Y S L K D L I T K V S K E S D N V A H N L L G Y Y I S N O S D A T 260

901 CATTCAAATCCAAGATGCTGCAATATCGGAGATGATCGGATCCAAAAGAAAAATTCATTTCTTCAAGATGCCCGGGAAGTTTATGGAAGCTATTTA 1000  
GTAAGTTTAGGTTCTACAGACGGTAATACCTTCTACTAACCTAGGTTTCTTTTAACTAAAGAGATTTCTACCGGCCCTTCAATACCTTCGATAAAT

261 F K S K M S A I M G D D N D P K E K L I S S K M A G R F M E A I Y 293

1001 TAATCAAAATGGATTTGTGCTAGAGTCTTTGACTAAACAGATTTTGATAGTCAGCGAATTCGCAAGGTGTTTCTGTTAAAGTAGCTCATAAATTTGGA 1100  
ATTAGTTTACCTAAACACGATCTCAGAACTGATTTGCTTAAACTATCAGTCCGTTAACGGTTCCCAAAAGACAAATTCATCGAGTATTTAACT

294 N C N G F V L E S L T K T D F D S O R I A K G V S V K V A N K I G 326

1101 GATCGGATGAATTTAAGCATGATACGGGTGTTGCTATGCGAGATTCCTCATTTATCTTTCTATTCTTCACTAAGAAATTCGATTATGATACGATTTCTA 1200  
CTACGCTTACTTAAATTCGTACTATGCCACAAAGATACGCTTAAGAGGTAAATAAAGAAAGATAAAAGTGATTTCTAAGACTAATACTATGCTAAGAGAT

327 D A D E F K N D T G V V Y A D S P F I L S I F T K N S D Y D T I S K 360

1201 AGATAGCCAGGATTTTATGAGGTTTAAATGAAGGAAACAGATTTTAAATCATTTCTCAAGAGGGATATTCTCAAAAGCATGCTAAGGCGGCTT 1300  
TCTATCGGTTCTTACAAATCTCAAGATTTTACTCCCTGGTCTAAAAATTTAGTAAAGAGTCTTCCCTATAAAGTTTTCGTACGATTCGCCCA

361 : A K C V Y E V L F \* 371



gsp76

Fig. 23

(SEQ ID NO: 68) 1 TTGAAAAATATTATCTATAAGAACGACATATAAATGTAAACAAAGGCGTAATATTATTAGGCCCTTTTTCGTATACTAGTATTGTCTTTAAAAGAAGGA 100  
(SEQ ID NO: 69) AACTTTTTATAATAGATATTCTTGTGTATATTACATTGTTCGCCATTATAAATAATCCGAAAAAACCATATGATCATAAACAGAAATTTTCTTCTCT

101 GTATCTACCTAATATGAAGAAAAAATCTTAGCGTCACTTTTATTAGTACAGTAATGGTTTCTCAAGTAGCTGTTTAACTGCGCATGCAGAAAGC 200  
CATAGATGCATTATACTTCTTTTGAATCGCAGTGAATAATTCATGTCAATACCAAGAGTTCATCGACAAAAATGTTGACCGCGTACGTCTTTGC

(SEQ ID NO: 67) 1 M K K K I L A S L L L S T V M V S Q V A V L T T A H A E T 29

201 ACTGATGACAAAATTCCTGCTCAAGATAATAAATTAGTAACTTAACAGCACAACCAAGAGCCCAAAAACAAGTTGACCAAAATTCAGGAGCAAGTAT 300  
TGACTACTGTTTTAACGACGAGTTCATTATTTTAATCATTGAATTGTGGTGTGTTCTTCGGGTTTTGTTCAACTGGTTTAAAGTCTCTGTTTATA

30 T D D K I A A Q D N K I S N L T A Q Q Q E A Q K Q V D O I Q E Q V S 63

301 CAGCTATTCAAGCTGAGCAGTCTAACTTGAAGCTGAAAAATGATAGATTACAGCAGAAATCTAAGAAAATCCAGCGCTGAGATTACAGAACTTTCTAAAAA 400  
GTCGATAAGTTGACTCGTCAGATTGAACGTTTCGACTTTTACTATCTAATGTTGCTCTTAGATTCTTTCAGCTCCCACTCTAATGCTTTGAAAGATTTTT

64 A I Q A E Q B N L Q A E N D R L Q A E S K K L E G E I T E L S K N 96

401 CATTTGTTCTCGTAACCAATCGTTGGA AAAACAAGCTCGTAGTGCTCAAA CAATGGAGCCGTAAGTATATCAATACCATTTGTAATCTCAAAATCA 500  
GTAACAAAGAGCATTGGTTAGCAACCTTTTGTTCGAGCATCAGCAGTTTGTTTACTCGGCATTGATCGATATAGTTATGTTAATCACTTTGAGTTTGTAGT

97 I V S R N Q S L E K Q A R S A Q T N G A V T S Y I N T I V N S K S 129

501 ATTACAGAAGCTATTTCACTGTTGCTGCAATGAGTGAATCGTATCTGCAAAACA AAAATGTTAGAACAACAAAGGCAGATAAAAAAGCTATTTCTG 600  
TAATGCTCTTCGATAAAGTGCAACGACGTTACTCACTTAGCATAGACGTTTGTGTTTTACAATCTTGTGTTTTCCGTCTATTTTTTCGATAAAGAC

130 I T E A I S R V A A H S E I V S A N N K M L E Q Q K A D K K A I S E 163

601 AAAAAAAGTAGCAATAATGCTATCACTACTGTAAATTCCTAATCAACAAAATTCGCTGATGATGCTCAAGCATTGACTACGAAACAGGCAGAACT 700  
TTTTTGTTCATCGTTTATTAAGTACGATAGTTATGACATTAACGATTAGTTGTTTTTAACCGACTACTACGAGTTCGTAATGATGCTTTGTCCTGCTTGA

164 K O V A N N D A I N T V I A N O O K L A D D A Q A L T T K O A E L 196

701 AAAAGCTGCTGAATTAAGTCTTGTGCTGAGAAAGCGACTAGCTGAAGGGGAAAAAGCAAGGCTATTAGAGCAAGAGCAGCAGCTGAGGCAGAGGCTCG 800  
TTTTCGACGACTTAATTCAGAACGACGACTCTTTCGCTGATCGACTTCCCCTTTTTCGTTCCGATAATCTCGTTCTTCGTGCTCGACTCCGTCTCCGAGC

197 K A A E L S L A A E K A T S \* 211

Fig. 24

YNES\_BACSV

(SEQ ID NO: 71) 1 ATGTTAATGCTTTATTGATTATTTGGCTACTTGTAGGCGAGCATTCCATCTGGCTTAATTGTGGCAAGCTTCCCAAAGGAATTGATATTCGGGAGC 100  
(SEQ ID NO: 72) TACAAATTAACGAATAACTAATAAAACCGATGAACCTATCCGTGTAAGGTAGACCGAATTAAACACCGTTGGAACGGTTTCCTTAACATAAGCCCTCG  
(SEQ ID NO: 70) 1 M L I A L L I I L A Y L I G S I P S G L I V G K L A K G I D I R E H 34

101 ACGGAAGCGGCAACTTAGCGCTACCAATGCATTCCGTACATTGGGTGTAAAAGCTGGTTCGGTGTCAAGCGGAGATATTTTGAAGGGACACTGCC 200  
TGCCTTCGCGTTGAATCCGCGATGGTTACGTAAAGCATGTAAACCACATTTTCGACCAAGCCAGCAGTATCGGCTCTATAAAACTTTCCCTGTGACCG  
35 G S G N L G A T N A F R T L O V K A G S V V I A G D I L K G T L A 67

201 AACTGCATTGCTTTTCTCATGTCATGTTGATATTCACCGCTTCTTGCAGGAGTCTTTGGCGTTTAGGCCACGTGTTTCCCATCTTCGCCAAATTTAAA 300  
TTGACGTAAACGAAAGAGTACGTACAACTATAAGTGGCGAAGAACGTCTCAGAAACGCCAAATCCGCTGCACAAAGGTAGAAGCGGTTTAAATTT  
68 T A L P F L M H V D I H P L L A G V F A V L G H V F P I F A K F X 100

301 GCGGTAAAGCGGTGGCCACATCAGGAGCGTTTTGCTATTTACGCAACCGCTTATTTATCAGCATGGTTGCGGTATTTCTCATCTTTTATACTTGA 400  
CGCCATTTTCGGCACCGCTGTAGTCTTCGCAAAACGATAAAATGCGTGGGACATAAATAGTCTACCAACGCCATAAGAAGTAGAAAAATATGAAC  
101 G G K A V A T S G G V L L P Y A P L L F I T H V A V F F I F L Y L T 134

401 CTAAATTTGTTTCTCTCATGATGTTAACAGGGATCTATACTGTTATATATAGTTTCTTTGTCATGATACGTATTTATGATTGTCCTTACCGTCT 500  
GATTTAAACAAAGAGAGTAGCTACAATTTGCTCCTAGATATGACAATATATATCAAGAAACAGGTACTATGCATAAATAACTAACAGCAATGGGACGA  
135 K F V S L S S M L T G I Y T V I Y S F F V H D T Y L L I V V T L L 167

501 CACTATTTTGTGATATACAGACACCGAGCGAACATTAACGAATTATCAATAAAACAGAACCTAAAGTAAATGTTATAA 582  
GTGATAAAACACTATATGTCGTGGCTCGCTGTAAATTTGCTTAATAGTTATTTGTCTTGGATTTCATTTTACCAATATT  
168 T I F V I Y R N R A N I K R I I N K T E P K V K M L \* 193

# Strategy for the targeted deletions of genes in *S. pneumoniae*

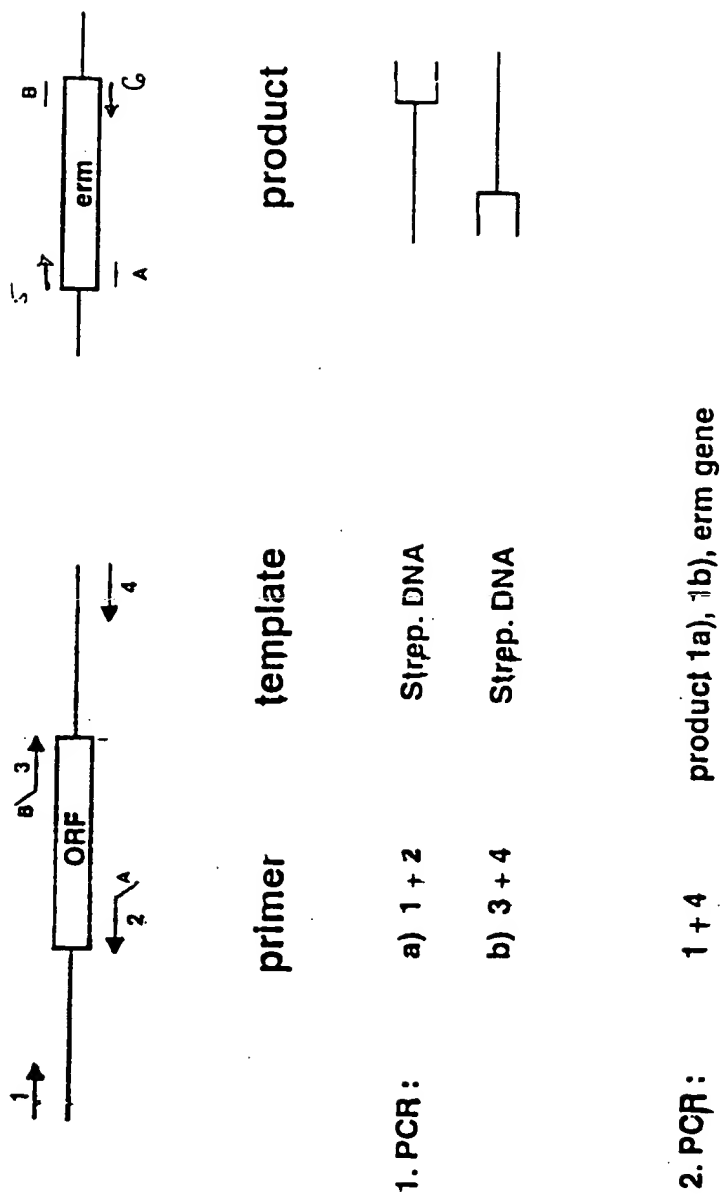


FIG. 25

Non-polar gene knockouts in *S. pneumoniae*

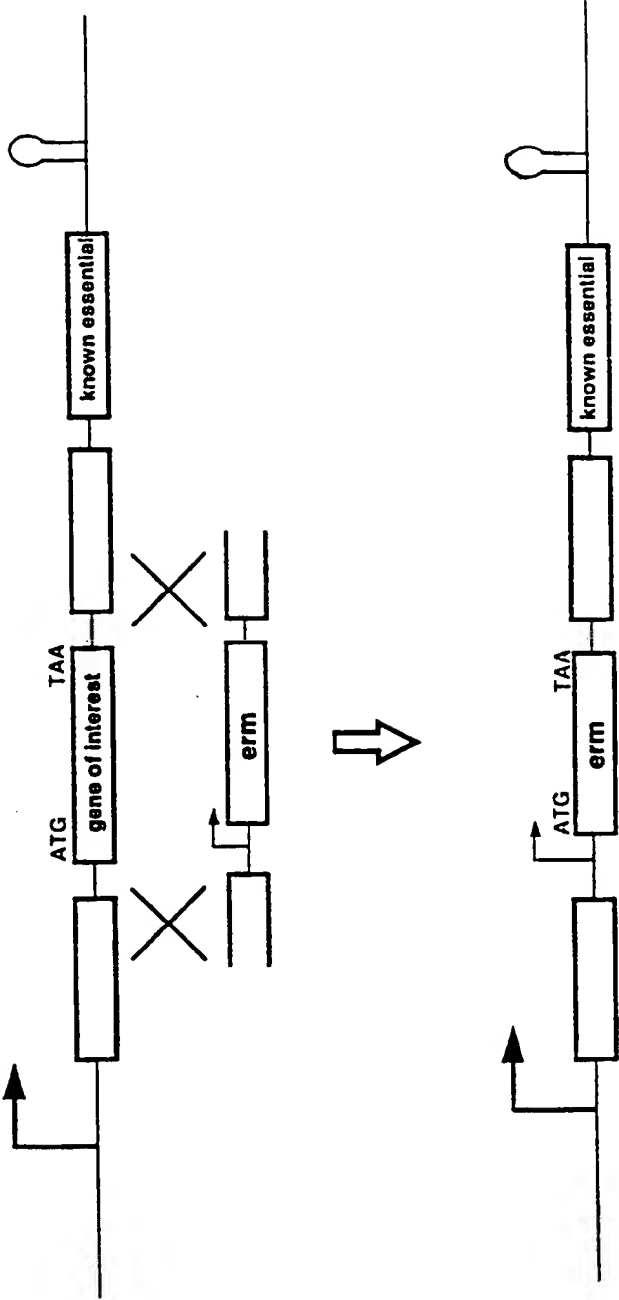


FIG. 26

- 1 -  
Sequence Listing

gcp103

(SEQ ID NO: 2) 1 TCCTGATTTTGGAGAAAGTTTATTAGAGATAAAAGAGCTAAGGAAAAAATTCATTTGATATTTTCTTCTATAAAATAGATAAAAAATGGTACAATA 100  
ACGACTAAAAACCTCTTTCAAAATAATCTCTATTTTCTCAGATTCCCTTTTAAAGGTAAACTATAAAAGAAGATATTTTATCTATTTTACCATGTTAT

(SEQ ID NO: 3)

101 ATAAATTGAGGTAATAAGGATGAGATTAGATAAATATTTAAAGTATCGCGAATTATCAAGCGTCGTACAGTCGCAAGGAAGTAGCAGATAAAGGTAGA 200  
TATTTAACTCCATTATTCCTACTCTAATCTATTTATAAATTTTCATAGCGCTTAATAGTTCGCAGCATGTCAGCGTTTCCTTCATCGTCTATTTCCATCT

(SEQ ID NO: 1) 1 M R L D K Y L K V S R I I K R R T V A K E V A D K G R 27

201 ATCAAGGTTAATGGAAATCTTGGCCAAAAGTTCAACGGACTTGAAAGTTAATGACCAAGTTCAAAATTCGCTTTGGCAATAAGTTGCTGCTTGTAAAAGTAC 300  
TAGTTTCAATTACCTTAGAACCGGTTTTCAAGTTGCTGAACTTTCAATTACTGGTTCAACTTTAAGCGAAACCGTTATTCAACGACGAACATTTTCATG

28 I K V N G I L A K S S T D L K V N D Q V E I R F G N K L L L V K V L 62

301 TAGAGATGAAGATAGTACAAAAAAGAGATGCAGCAGCAATGTATGAAATTATCAGTGAACACGGGTAGAAGAAAAATGCTAAAAATATTCTACAAT 400  
ATCTCTACTTTCTATCATGTTTTTCTTCTACGTGCTCCTTACATACTTTAATAGTCACTTTGTGCCCATCTTCTTTTACAGATTTTATAACATGTTA

62 E M K D S T K X E D A A G H Y E I I S E T R V E E N V \* 89

gcp1119

(SEQ ID NO: 5) : GAAATCCCTTTCCAAATGTGACTGTACCCATGAACCGCTTTATGAACGCTTTTCCAGCTTTCCAGCTTCAGACTTACAGGAAATGAAAGAGGAGACCAACG 100  
(SEQ ID NO: 6) : CTTTAGCCAAAGGTTACACTGACATCGGTACTTGGGAAATATCTTGGAGAACGGTCCGAAGGTTCCAGTCTGAATGTCCTTTACTTTCTCTCTCTGGTGC

101: GGGCAGAAATCACTTGTCAATTCTGCCAACTACTTACAACTTTGATCAAAAGGACCTGCAGGAATCAATTCTGCACAAATCTTAATAACCTTTTATCA 200  
CCGCTCTTACTGAACAGTTAAGACGGTTGATGAATGTTGAAGCTATCTTTCTCGACCTCTTGAATAGCACTGTTTAGAATTTATGCGAAATATCT

(SEQ ID NO: 4) : M K R T M R N S F V T N L N T P F M I 19

201: TTGGCAATATTGAGATTCCCAATCTACCTTTAGCGCTTATGGCTGGCTGACCAACTCAGCTTTGGTACCATGCGAAAGAGCTCGGAGCTGCACT 300  
AACCTTTATAACTCTAAGGCTTAGCATGGCAAAATCGCGGATACCGACCGCACTGGTTGAGTCCGAAAGCATGGTAGCGTTTCTCGAGCTTCGACCTGA

20 G N I E I P N R T V L A P M A G V T N S A F R T I A K E L G A G L 52

301: CTTGTATATGAAATGCTCTGTGACAGGGAAATCCAAATACAAACGAAAAAACCTTGCAATGCTTCATATCGATGAGGGCGAAAAACCTGTCTCTATC 400  
GCAACATTACTTTACGAGACTCTTCTCTTAGGTTATGTTGCTTTTGGGACGTATACGAAGTATAGCTACTCCCGCTTTGGGACAGAGATAG

53 V V N E M V S D K G : Q Y N N E X T L H M L H I D E G E N P V S : 85

401: CAATTTTGGTACCGATCAAGACAGCTTACCACTGGCGAGCAGAAATTCATCAAGAAAAACCAAGACCGATATCGTCAATATCAACATGGGCTGCCCTG 500  
GTTGAAAAACCATCGCTACTTCTGCGATCTGGCGCTCTCTTAAGTAGGTTCTTTGTTGTTCTGGCTATAGCGCTATAGTTGTACCGACGGGAC

86 C L F G S D E D S L A R A A E F : Q E N T X T D I V D I N H G C P V 119

501: TCAACAAATCTGCAAGAACGAACCTGAGCTATGTGGCTCAAGATCTGACAAAGTCTACTCTATCATCAACAGGTCAGTCTGTCTTGATATCCC 600  
AGTTGTTTAGCACTTCTTCTTCCAGCTCGATACCGAGTTCTTAGCACTGTTCTAGATGAGATAGTGTGTTCCAGGTCAGACAGGAATATAGGG

120 N K : V K N E A G A M M L K D P D X I Y S I I N K V Q S V L D I P 152

601: ACTTACTGTCAAAATGCTACCGGCTGGGCGGACCATCTTGGCAGTAGAAAAATGCCCTCGCTGAGGCTGCAGGTGTTTCTGCCCTCGCCATGGCAT 700  
TGAATGACAGTTTACGCACTGGCCGACCGCTGGGTAGACACCTCTCTTTACGGGAGCGGAGCTCCGACGTCACAAAGACGGGAGCGGTACGTA

153 L T V K M R T C W A D P S L A V E N A L A A E A A G V S A L A M H 185

701: GGGCGTACCGCTGAACAAATGTATATGCGCCACCGACACTTGACACCTTTACAAGGTTGCCCAAGCTTAACCAAGATTCCATTTCATCGCCAAACGGTG 800  
CCGGCATGGGCACTGTTTACATATGACCGGTGCTGTGGAATCTTGGAAATGTTCCAACGGGTCGAGATTGTTCTAAGGTAAAGTAGCGGTGGCAC

186 G R T R E O M Y T G H A D L E T L Y K V A O A L T K I P F I A N G D 219

801: ATATCCGTACTGTCGAAGAGCCAAAGCAACCGATCGAAGAAAGTTGGTGTGACGCACTCATGATTGGCCGAGCTGCCATGGGAAATCCTTACCTCTTCAA 900  
TATAGGCATGACAGGTTCTTCTGCTTCTGGTAGCTTCTTCAACCAAGACTGCGTCAGTACTAACCGGCTCGACGGTACCCTTTAGGAATCGGAGAGTT

220 : R T V Q E A K C R I E E V G A D A V M I G R A A M G N P Y L F N 252

901: CCAATCAACCATTACTTTGAAACAGGAGAAATCCTACCTGATTGACCTTTGAAGACAAGATGAAGATCGCCACGAACACTTGAACAGGATTGATTAC 1000  
CGTTTAAATGGTAATGAACTTTGCTCTTTAGGATGGACTAAACTTGGAACTTCTGTTCTACTTCTAGCGGATGCTTGTGAACCTTGTCTAACTAATTG

253 Q I N H Y F E T G E I L P D L T F E D K M X I A Y E H L K R L I N 285

1001: CTCNAAGCAGAAAACTCCCACTTCTGAAATCCCGCGCTCGCTCTCTACTATCTCCGCAACATCTGGCGCTGCCAACTCCGTGAGCCATTTCCG 1100  
GAGTTTCTCTTTTCAGCGCTCAAGCACTTAAGCGCGCGAGCGAGGAGTGATAGAGGCACCTTGTAGACCGGACGTTTGAAGCACTCGTAAAGCG

286 L K G E N V A V R E F R G L A P H Y L R G T S G A A K L R G A I S O 319

1101: AAGCTAGCACCTTAGCAGAGATTGAAGCCCTCTTGCAATTGGAGAGGCTTAATAGTTTAAAAACCGTAACCTCTCTTAAGAGTCTCTTGAATGCCGCCA 1200  
TTCCATCGTGGATCTCTCTAACTTCCGAGAACGTTAACTCTTCCGAATTATCAAAATTTGGGCATTGAGAGAATTTCTCAGAGAACTTACGGCGCT

320 A S T L A E I E A L L O L E R A \* 336



249 S E I Y G E K T E : I W G O E S I O E G V L N Y E F S L S P R A F 281  
1401 TTATCAACTAAATCCTGAGCAACAGAACTCTCTATAGCGAAGCAGTAAAGCGCTGGATGTTTCATAAGAAGACCATTTGATTGACGCTTATTGTGGA 1500  
AATAGTTGATTTAGCACTCTTTCTCTTCAGGAGATATCGCTTCCTCATTTTCGGACCTACAACTATTTCTTCTGGTAAACTAACTGCGAATAACACCT  
282 Y O L W P E O T E V L Y S E A V K A L D V D K E D N L I D A Y C G 314  
1501 GTTGGAACTGATTGCGATTTGCTTTCAGAAAGTAAAACTCAGAGGTATGGATATTATTCAGAAAGCTATTGAAGATGCCAAGCGAAATGCTAAAA 1600  
CAACCTTGCTAACCTAAACGGAACGCTTCTCTCACTTTGTGAGTCTCCATACCTATAATAAGGTCTTCGATAACTTCTACGCTTCGCTTACGATTTT  
315 V G T I G F A F A K K V K T L R G M D I I P E A I E D A K R N A X R 348  
1601 GAATGGGATTTGACAACTACTCATTATGAAGCTCGAAGCGCAGAGAGATTATTCCTCTTGGTACAAGGAAGGCTACCGAGCAGATGCTTGTATTGTA 1700  
CTTACCTCTAAACTGTTATGAGTAATCTTCGACCTTGGCTCTTCTCTAATAAGGAGCAACCATGTTCTTCCGATGGCTCTCTACGAACTAACCACT  
349 M G F D N T H Y E A G T A E E I I P R W Y K E G Y R A D A L I V D 381  
1701 CCCACCACTACAGCTCTGGATGATAAGTTATTAGATACTATTCTTACTTATGTACCAGAAAAATGGTTTATATTTCTTGTAAATGTTTCGACCTTGGCT 1800  
GGGTGGTGGATGTCAGACCTACTATTCAATAATCTATGATAAGAAATGAATACATGGTCTTTTTTACCATAATAAAGAACATTACAAAGCTGGAACCGA  
382 P P R T G L D D K L L D T : L T Y V P E K M V Y I S C H V S T L A 414  
1801 CCGTATTGGTACGCTTAGTAGAAGTCTATGATCTTCATTATATCCAGTGGGTGGATATGTTCCACATACAGCTCGAAGTGAAGCTGTTGTAAATTTAA 1900  
CCACTAAACCAATGCGAATCATCTTCAGATACTAGAAGTAATATAGGTCAAGCCAGCTATACAAGGGTGTATGTTCGAGCTTGACTTCGACCAACATTTTAAT  
415 R D L V R L V E V Y C L H Y I O S V D M F P H T A R T E A V V K L I 448  
1901 TAACAAAAGTTTAAAAAAGTACTTCACAACTTTGAAAAGCTGTATATAAGTAAGAGTTGAAAAATAACAACCTCAGGTNCGTTGCTCAAGGGGTTAAGAC 2000  
ATTGTTTTCAAAATTTTTCATCAACTGTTTCAAACTTTTTCACATATTATCATTTCTCAACTTTTATTGTTGAGTCCAGCAACCACTTCCCCAATTCTG  
449 T K V . 452  
2001 AGCGCTTTTCAAGCGCTTAACAGGGGTTGAAATCCCGTACGGACTATGGTATGTTGGCGTTGGCAACTTGTATGAAAACTTTA 2084  
TCCGAAAAGTGCCGCAATTGTGCCAAGCTTAGGGCATGGCTGATACCATACAAAGCCAACTTGTGAACACTTTTGAAT



(SEQ ID NO:11)	1	106
(SEQ ID NO:12)		
	101	200
(SEQ ID NO:10)	1	15
	201	300
	16	49
	301	400
	50	82
	491	500
	83	115
	501	600
	116	149
	601	700
	150	182
	701	800
	183	215
	801	900
	216	226
	901	1000

gcp1493

(SEQ ID NO:14) : TAAAGACACTGGAACGACCAACACCTTCCGATTTAGGTAAGAAAGCTGGTATGGCAACCTTGTGATTGACTTTTCAAGGAACCTAGCAACGCTG 100  
(SEQ ID NO:15) : ATTTCTGTGACCTTCTGCTTGTGGAAGGCGTAAAAATCCATTCTTTGCACATAACCTTGGAAACACTAACTGAAAAAGTTTCTTTGGGATGCTTCCGAC  
(SEQ ID NO:13) 1 K D T G T T N T F R I L G K K A G H A T F V I D F F K G T L A T L 33

101 CTTCCGATTATTTT CATCTACAAGGCGTTCTCTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTGCAAGGATTTAAAGGTG 200  
GAAGGCTAATAAAAAGTAGATGTTCCGCAAGAGGAGAGTAGAAAACCTGAAAACCGACAATAGCCGCTATGGAAGGGATAGAAAACGTCCTAAATTTCCAC  
34 L P I I F H L Q G V S P L I F G L L A V I G H T F P I F A G F K G G 67

201 GTAAGGCTGTGCAACCACTGCTGGAGTGATTTCCGATTTCGGCTATCTTCTGCTCTACCTTGGGATTATCTTCTTTGGACTCTCATATCTTGGCAG 300  
CATTCGGACAGCGTTGGTCACGACCTCACTAAAAGCCTAAACGCGGATAGAAGACAGAGATGCAACGCTAATAGAAGAAAACCTGAGAGTATAGAACCGTC  
68 K A V A T S A G V : F G P A P I F C L Y L A : I F F G L S Y L G S 100

301 TATGATTTCACTGTCTAGTGTCAAGCATCGATCCCGCTGTTA 344  
ATACTAAAAGTCACAGATCACAGTGTCTAGCTAGCGCGACAAT  
101 M I S L S S V T A S I A A V 114

gcp150:

(SEQ ID NO:17) : CTAAAGGTAAATTCGAATGAAAAGTATAAAATTAAATGCTCTATCTTACATGGGAATTCCTCTTGAATATTATTTCCCATCCTAACTGGAACCTATG 100  
(SEQ ID NO:18) : GATTTTCATTTAACTTACTTTTCATATTTAAATTTACGAGATAGAAATGTACCTTTAAGCACAGAACTTATAATAAAGGGTAGGATTGACCTTGGATAC  
(SEQ ID NO:16) : M K S I K L N A L S Y M G I R V L N I I F P I L T G T Y V 29

101 TCGCGGTGCTCTTGGACCCAACTGACTATGGTTACTTCAACTCAGTCGACACTATTTGTCATTTCTTGGCTTTGCAACTTATGGTGTCTATAACTA 200  
AGCCGGCAGAGAACCTGGCTTCACTGATACCAATGAAGTTGAGTCAGCTGTGATAAAACAGTAAAAAGAACGGGAAACGTTGAATACCACAGATATTGAT  
30 A R V L D R T D Y G Y F N S V D T I L S F F L P F A T Y G V Y N Y 62

201 CGGTTTAAGGGCTATCAGTAATGTCAGGATAACAAAAAGATCTTAACAGAACCTTTCTAGTCTTTTATTGTGCAATGGCTGTACCATTTTGACC 300  
GCCAAATTCGGATAGTCATTACAGTTCTTATTTTCTAGAAATGCTCTGAAAAGATCAGAAAAATAAACAGTACCGAACATGCTAAAGCTGG  
63 G L R A I S N V K D M K K D L N R T F S S L F Y L C I A C T I L T 95

101 ACTGCTGTCTATATCTTACCTATCTCTCTCTTTACTGATAATCCAATCGTCAAAAAGGTCTACCTTGTATGGGGATTCAACTCATTGCCAGATT 400  
TGACGACAGATATAGGATCGGATAGGAGAGAGAAATGACTATTAGGTTAGCAGTTTTCCAGATGGAAACAATACCCCTAAGTTGAGTAACGGGTCTAAA  
96 T A V Y I L A Y P L F F T D N P I V K K V Y L V M G I Q L I A Q I F 129

401 TTTCAATCGAATCGGTCAATGAAGCTTGGAAAATTACAGTTTCTCTTTTACAAAACCTGC 460  
AAAGTTAGCTTACCCAGTTACTTGGAGACTTTTAAATGTCAAAGAGAAAAATGTTTTGACG  
130 S I E W V N E A L E N Y S F S F T R L 148

gcp151:

SEQ ID NO: 20) : CCGCCATTTACCGTGAAGATTTCAGCTATGTAATGATTTTATGCAACACGTCAGAGCAGGACGGAATGATGTTTGTGACGAGTTGCTATACA 100  
GCAGCGTAAATGGCACTACTAAAGTCATACCTACTAAAAATACCTGTTGCGAGCTCTCGTCTGCTCTTACATACAAAACACTGCTCAACGATATGT

SEQ ID NO: 21) 101: GCGAGTAGGCAATGCAGATTCAAAAAAGTTTTAAGGGCAGTCTCCCTATGGCAAGCTGTATCTAGTGGCAACGCGGATTGGCAATCTAGATGATGACT 200  
CCCTCATCGGTACGTCTAAGTTTTCAAAAATCCCGTTCAGAGGGATACCGTTCGACATAGATCACCGTTGCGGCTAACCGTTAGATCTACTATACTGA

SEQ ID NO: 19) 1 M O I O K S F K G O S P Y G K L Y L V A T P I G N L D D N Y 30

201: TTTCTGCTATCCAGACCTTGAAGAAGTGGACTGGATTGCTGCTGAGGATACGCGCAATACAGGGCTTTTGTCTCAAGCATTITGACATTTCCACCAAGC 300  
AAAGCAGATAGGTCTGGAACTTTCTTCACTGACCTAACGAGGACTCTATGCGGTTATGTCCCGAAAACGAGTTTCGTAAACCTGTAAGGTGGTTCTG

31 F R A I O T L K E V D W I A A E D T R N T G L L L K H F D I S T K Q 64

321: AGATCAGTTTTCATGAGCACAATGCAGAGGAAAAATTCCTCATTTGATTGGTTCTTGAAGCAGGCGAAAGTATGCTCAGGCTCTCTGATGCCCGTTT 400  
TCTAGTCAAAAGTACTCGTTTACGTTTCTTTTTAAGGACTAACTAACCAAGAACTTTCTGCCGTTTCATAACGAGTCCAGAGACTACGGCCAAA

65 I S F H E H N A K E K : P D L I G F L K A G O S I A O V S D A G L 97

401: GCGTAGCATTTGAGACCTGTCATGATTAGTTAAGGCAAGTATTGAGGAAGAAATTCAGTTGTGACTGTTCCAGGTACCTCTGAGCAATTTCTGCC 500  
CGGATCGTAAAGTCTGGACACGTAATAATCAATTCGTGATAACTCTTTCTTAACTCAACACTGACAAGGTCATGGAGACGTCCTTAAAGACGG

98 P S I S D P G H D L V K A A I E E E I A V V T V P G T S A G I S A 130

501: TTGATTGCCAGTGGTTAGCGCCACAGCCACATATCTTTACGGTTTTTACCGAGAAAATCAGGTCAACAGAGCAATTTTTTGGCTCTAAAAAAGATT 600  
AACTAAAGGTCAACCAATCCCGGTGTCGGTGTATAGAAAATGCCAAAAAATGGCTCTTTTAGTCCAGTTGTCTTGGTTAAAAAACGAGATTTTTTCTAA

131 L I A S G L A P C P H : F Y C F L P R K S G O Q K O F F G S K K D Y 164

601: ATCTGAAACACAGATTTTATGAATCACTCATCTGTAGCAGACAGTTGGAATAATGTTAGAAGTCTACGGTGACCGCTCGGTGTTTGTCTCAG 700  
TAGGACTTTGTGCTAAAAAATACTTAGTGGATAGCAGATCTCTGTGCAACCTTTTATACAATCTTCAGATGCCACTGGCGGCAACAAAAACAGTC

165 P E T Q I F Y E S P H R V A D T L E N M L E V Y G D R S V V L V R 197

701: GGAATTGACCAAAATCTATGAAGAATACCAAGAGGTACAATTTCTGAATGCTGGAAGCATCTCTGAAACGTCCTCAAGGGTGAATGTTCTTGATT 800  
CTTAACTGGTTTTAGATACTTCTTATGTTTCTCATGTTAAAGACTTAACGACCTTTCTAGAGACTTTGACAGAGTTCCCACTTACAGAGACTAA

198 E L T K I Y E E Y O R C T I S E L L E S I S E T S L K G E C L L I 230

801: GTTGAAGGTGCCAGCAAGGTGTGAGGAAAAAGSAGGAAAGACTTTGTTCTTACAAATCCAAGCCCTATCCAGCAAGGCATGAACAAAAATCAAGCTA 900  
CAACTTCCAGGTGTTTCCACACCTCTTTCTACTCTTCTGAAGCAAGATCTTTAGGTTCCGGCATAGGTGTTCCGTACTTCTTTTATGTCAT

231 V E G A S K G V E E K D E E D L F L E I O A R I O O G M K K N Q A I 264

901: TTAAGCAATAGCTAAGATTACCACTGCAATAAGAGTCACTCTACGCTGCTTACCAGCACTCGGAAGAAAAAATAAAGGGACACAGGATGTAATAA 1000  
AATTCCTTTATCGATTCTAAATGGTCACTTATCTCAGTTGAGATGCCAGGATGGTGACCCCTTCTTTTGTATTTCCTCTGCTACATTATT

265 K E I A K I Y O W N K S O L Y A A Y N D M E E K Q \* 290

gcp1518

- 9 -

(SEQ ID NO: 23) 1 ATGGCTTGGTTAAAAAAGGTGGCAATGCTCTTTAAGTGCAGTTATTGCGCTGTAGCATATAAATCTATTCTACATATTTTAAACGTTCTACGAC 100  
(SEQ ID NO: 24) TACCGAACCAATTTTTCACCGTTACGAGAAATTCAGTTCAATAACGGACATGATATTTAGATAAAGGATGTATAAAAAATTTGCAAGATGCTC

101 TTAATTTGAAACGTTTAGCTTGTGGTATAATAGATTTATGGATAAAAAATATGAAAAATCTCTCAGGATTGGGAATGACGTTAAAGCAATTGATACC 200  
AATTAACCTTTGCAATCGAACACCATATTATCTAAATACCTATTTTATACCTTTTAGAGAGTCTTAAACCTTCACTGCAATTTGTTTAACTATGG

(SEQ ID NO: 22) 1 M D K K Y E K I S Q D L G V T L K Q I D T 21

201 GTTCTAAGTTTGACAGCTGAAGGGGGGACTATTCCCTTTATCGCGGTTATCGCAAGGACATGACTGGTACTGTGGATGAGGTCGCCATTAAAGCTATTA 300  
CAAGATTCAAACTGTGACTTCCCGCTGATAAGGGAATAGCGCGCAATAGCGTTCTGTACTGACCATCAGAACTACTCCACCGCTAATTCGGATAAT

22 V L S L T A E G A T I P F I A R Y R K D M T G S L D E V A I K A I I 55

301 TTGATTTGGATAAAAGTGTGACAAATCTCAATGACCGTAAGCAAGCTGTCTTAGCTAGGATTCAAGAACAAAGTAAAGTTGACCAAGCAATTCGAAGAAC 400  
AACTAAACCTATTTTACAGCTGTTTAGAGTTACTGGCATTCCTTCGACAGAAATCGATTCTAAGTTCTTGTTCATTCAGCTGTTCTTAACTCTTCTG

56 D L D K S L T H L N D R K E A V L A K I O E O G K L T K E L E E A 88

401 TATCTTAGTTTCCGAAAAATTAGCAGACGTTGAAGAACTCTATCTTCTTATAAGGAAAGCGTCTTACCAAGGCAACCAATTCGCCGTGAAGCTGGACTC 500  
ATAGAAATCAACGGCTTTTAACTCTGTGCAACTCTTTCAGATAGAAAGAAATATCTCTTTTCGACGATGTTTCCGTTTGGTAAACGGGCACTTCGACCTGAG

89 : L V A E K L A D V E E L Y L P Y K E K R R T K A T I A R E A G L 121

501 TTTCTCTTGGCTGTTTGAATTTGCAAGATATAGTTGACTTAGAGAAAGAGCTGAAAGTTCTGTCTGGAAGGATTGCGACTGGCAAGGAAGCCTTGA 600  
AAAGGAGAACGAGCAAACTAAAGCTTTATATCAACTGAATCTCTTCTTCGACTTTTCAAGCAGACACTTCTTAAACGCTGACCTTCTTCGGAAT

122 F P L A R L I L O N : V D L E K E A E K F V C E G F A T G K E A L T 155

601 CCGGTGCAGTTGATATTTGGTCCAAAGCTTATCCGAAGATGTGACCTTTCGTTCTATGACTTATCAGGAAGTGTGAGACACTCTAAACTCACTTCTCA 700  
GGCCACCTCAACTATAAAACAGCTTCGAAATAGCTTCTACACTGGAGCGCAGATAGTGAATAGTCTTTCAGGACTCTGTGAGATTTCAGTGAAGAGT

156 G A V D I L V E A L S E D V T L R S M T Y O E V L R H S K L T S Q 188

701 AGCCAAAGATGAAACTCTTGATGAAAGCAGCTTTTCAGATTTATTATGATTTTCAGAGACAGTTGGAAGCTATGCAAGGCTATCTGACTTGGCTCTC 800  
TGGTTCTTACTTTTCAAGACTACTTTTCTGTCGAAAGCTTAAATAATATTAAGAGTCTCTGTCAACCTTGATACGTTCCGATAGGATGGAACCGAGAG

189 A K D E S L D E K O V F O : Y Y D F S E T V G T M O G Y R T L A L 221

801 AATCGTGGGAGAACTTGGTCTTGAAGATCGGTTTGAACATGCGACCGACCGTATTCTTGCCTTCTTTGGTACTGTTTCAAGCTGAAAAATGCTT 900  
TTAGCACCTCTCTTGAACACAGAACTTACGCAAACTTGTACGCTGCTGGCATAAGAACGGAAGAACGATGAGCAAGTTCCACTTTTACGAA

222 N R G E K L G V L K : G F E H A T D R I L A F F A T R F K V K N A Y 255

901 ATATTGATGAAGTTGTTGAGCAATCCGTTAAGAAAAAGGTCCTGCTGCTATTGAGCGTCTGATTCCGACAGAAATTAAGTGAGAAAGCTGAAGAGGGAGC 1000  
TATAACTACTTCAACAAAGTCGTTAGGCAATTTTTCGAGAACCGACGATAACTCGCAGCATAGCGCTGCTTAATTGACTCTTTCGACTTCTCCCTCG

256 : D E V V O O S V K K K V L P A I E R R I R T E L T E K A E E G A 288

1001 TATCCAACTTTTTCGCAATCTCGCAATCTCTCTTCTGCTGCTCACTGAAAGGGCGCGTGGTTCTTGGATTGACCCGACCTTCTGTACAGGTGCC 1100  
ATAGGTTGAAAAAGACTGTTAGACGCTTAGAGGAGAACCAACGAGGTGACTTTCCCGCGCACCAAGAACTAAACTGGGTGGAAGCATGTCCACGG

289 : O L F S D N L R N L L L V A P L K G R V V L G F D P A F R T G A 321

1101 AAGTTAGCTGTCTGATGCAACAGGAAAAATGCTGACAACTCAGGTTATTTATCTGTTAAACCAAGCATCAGCTCTCAATCGAAGAACCAAGAAAG 1200  
TTCAATCGACAGCACCTACGTTGCTTTTACGACTGTTGAGTCCATAAATAGGACAAATTTGCTGATGCGAGCAGTTAGCTTCTTGGTTCTTTC

122 K L A V V D A T G K M L T T O V : Y P V K P A S A R Q I E E A K K D 355

1201 ATTTAGCAGATTAATGCTCAATACGCTGTAGAGATTATGCCATTGGAATGGAACCGCCAGTCTGCAAGTGAAGCTTTGTAGCGGAAGTTCTGAA 1300  
TAAATCGTCTAAATTAACGAGTTATGCCACATCTTAATAACGGTAACTTACCTTGGCGTCAGCACTTTCACCTCGAAGACATCGCCTTCAAGACTT

356 L A D L I G O Y G V E I I A I G N G T A S R E S E A F V A E V L K 388

- 10 -

1301 AGATTTCCTGGAAGTCAGCTATGTTATGCTTAATGAAAGTGGCTCTCTCTATTTCTGCCAGCGAATTCCTCTCAAGAGTTCCAGACTTCACCTT 1400  
TCTAAAGGGACTTCAGTCGATACAAATAGCAATTACTTTCACTACCAAGACAGATAAGACGCTCTTCAACGAGCAGTCTCTCAAGGCTCTGAAGTGGCAA

1389 D F P E V S Y V I V N E S G A S V Y S A S E L A R O E F P D L T V 421

1401 GAAAAACGCTCTGCCATTCTATGCCCTCTCTCTGCAAGATGCTCTTCCGCAATTGGTCAAAATCGATCTTAAGTCAATTGGTCTCGTCAATACCAAC 1500  
CTTTTCCGAGACGCTAAGATAGCGGGCAGCAACGCTCTAGGAGAACGCTTAACCACTTTAGCTAGGATTCAGTTAACCAAGCCAGTTATGGTTG

1422 E K R S A I S I A R R L O D P L A E L V K I D P K S I G V G Q Y Q H 455

1501 ACCATGTCAGTCAGAAGAACTATCTCAGAGCTCTGCACTTTGTTCTGATACAGTGGTTAACCAAGTTGGTCTCAATGTCATATACAGCTAGCCGAGCTT 1600  
TGCTACAGTCACTCTTCTTGTATAGACTCTCAGACCTGAAACACAGCTATGTCACCAATTGGTTCAACCACAGTTACAGTTATGTCGATCCGGTGGAGA

1456 D V S O K K L E E S L D F V V D T V V N Q V G V N V M T A S P A L 488

1601 TCTTTCACAGTAGCTGCACTCAACAAAATCTCTCTGAAAATATTGTCAATACCGCGAGCAAGCAAGCAAAATCACTTCACTCCGCCCCAAATCAAGAAA 1700  
AGAAAGTGTGCATCGACCTGAGTCTCTTGTATAGAGACTTTTATAACAGTTTATGGCGCTCTCTCTCTTTTATGTAAGTGGCGGCTTATGTTCTT

1489 L S H V A G L N K T I S E N I V K Y R E E E G K I T S R A Q I K K 521

1701 GTTCTCTCTCTGGAGCCCAAGGCTTTGAGCAGGCTGCTGGTTCTCTCTCTATCCCTGAAAGTAGCAATATCCTTGATATACAGGAGTTCACTCAGAG 1799  
CAAGGAGCAGACCTCTGCTTCCGCAACTCTGTCGACCAAGCAAGCATAGGGACTTTCACTGTTATAGGAACTATTATGTCCTCAAGTGGGCTCT

1522 V P R L C A K A F E O A A G P L R I P E S S N I L D N T G V H P E 554

g9p1346

- 11 -

(SEQ ID NO: 26) 1 TACTGGGCAAGGCTTCTACCCCTGTTGAAATGTAAGGCTTTCTTGAATAAGTGAAGTTAAGATTTTCAGAGCACTCAACGAAGCCAGNATCTCC 100  
(SEQ ID NO: 27) 1 ATGACCCCTCTTCCAAAGAAATGGGACAAGACTTACACTTCCAGAAAGAACTTTACCACTTCAATTTCTAAAGTCTCTGAGTTGCTTCTGCTGTTAGGCG 100  
(SEQ ID NO: 25) 1 T G A R V S Y P V L N V R V P L E N G E V K I P R A L N E A Z I R 113

101 AGGCTGATGCAACCAATGCTGCCAGATATTGTAAATAGTGTCTCTTGAACGTTTCTGAGACGCGCTAAGCTTTCAGACCGACTGGTACTA 200  
102 TCCAGACTAGCTTGGTACCACTCTATACATTATTACCAAGGGAACTTGCAGAACCACTTCTGCCCATTTGCAAGCTGTGGCTGACCATCAT 200  
103 R S D R T M V A D I V I N C V P P E R R F R G D G L T V S T P T G S T 67

201 CTGCTATAACAAGTCTCTTGGCTGCTCTTTACACCTACCAATTGAAGCTTTGCAATTACCGAGATTGCCAGCTTAATATCTGTCTATGGAAC 300  
202 GACCGATTATTGTCAGAGAACCCCAAGCAAAATGGCATGGTAACCTGAAACCTTAATTGCTCTAACCGTGGAAATTATTAGCAGATAGCTTG 300  
203 A V N K S L G C A V L N P T I E A L O L T E I A S L N N R V Y R T 100

301 ATTGGCTCTTCCATTATTGTCCTAAGAAGGATAAGATTGAATTTTCCAAACAAGAACTATTATCACTATTTCCGTTGACAATAGCTTTATTCT 400  
302 TAACCCGAGAAGGTAATAACACCGATTCTTCTATTCTAATCTGAATAAGGTTGTTCTTCTTAATAGTATGATAAGCCAACTGTTATCCCAATAGA 400  
303 L G S S I I V P K K D K I E L I P T R M D Y N T I S V D N S V Y S 113

401 TTCCTAATATTGAGCTATTGAGTATCAATCGACCATCATAGATTCACTTTCTCGGACTCTAGCCATACCACTTTCTGCAACGCTGTTAAGGATG 500  
402 AAGCGATTATACTCGCATAACTCATAGTTAGCTGGTAGTATTCTAAGTGAAAACAGCGCTGAGGATCGGTATGCTCAAGACCTTGGCACAATCTCTAC 500  
403 F R N I E R : E Y Q : D N N K I N P V A T P S R T S P W N R V K D A 167

501 CCTTTATCGGTGAAGTGGATGAATGAGCTTGAATTTATCCAGATGAACATGTCAAGGTTAAGACCTTTTAAAAAA 578  
502 GGAATAGCCACTCCACTACTTATCCAAACTTAAATAGCTCTACTTGTACAGTTCCAAATCTGGAATAATTTTTT 578

168 F I G E V D E . 175

gcp1551

- 12 -

(SEQ ID NO: 29): GGCTCTAAAAGAAACCTACTGCGAGCTGATAGATGGGAAGTACTATTAATTCCTCTATCCGAGAGATGGTTGTCCGCTGCCAATATATACCTGCT 100  
(SEQ ID NO: 30): CGAGATTTCTTTGGATGACCTCTCACTATCTACCTCTCATGATAATAAACTAGGAAATAGGCTCTCTACCAACAGCCGACCTTATATATGGAACCA 100  
(SEQ ID NO: 28): M V V G H Q Y I P A 10

101 CCACACAGGGGGTTACGATTGGTCTCTCCCAAGATAGACATTGCTCTTAGACCAGATTGGTTTATTTTGGTCAAGATGGTCTCTACAAGATTTC 200  
GGTGTCTCCCAATGCTAACCCAGGAAGAGCTTCTTATCTCTAACGAGAAATCTGGTCTAACCAAAAATAAAACCACTCTACCAACAGAAATGTTCTTAAC 200

11: P H K G V T : G P S P R I E I A L R P D M Y Y F G O D G V L O E P V 44

201 TTGGCAAGCAAGTTTATAGAAAGCAAAATGCTACGAATACCAACAAATCATGGGAAGAAATATGATAGCCAAAGCAGAGAAACGAGTCTTATTTTGA 300  
AACCGTTGCTTCAAAATCTTCTTTTGGAGATGCTTATGGTTGTGTAGTAGCCCTCTTATACTATCGGTTGCTCTCTTGGCTCAGATAATAAACT 300

45: G R O V L E A K T A T M T M E N H G E E Y D S Q A E R V Y Y F E 77

301 AGATCAGCGTACTTATCATACTTTAAAACCTGTTGCATTTATGAAGAGGGTTATGGTATTATTATACAGAAAGATGGTGGCTTTCATTTCCGATCAAC 400  
TCTAGTCCGATCAATAGTATGAAATTTTACCAACCTAAATATCTCTCCCAATAACCAATAAATGTCCTCTACCAACCAACTAAGAGCGTAGTTG 400

78: D Q R S Y H T L K T G M I Y E E G Y W Y Y L O K D G G F D B R I N 110

401 AGATTGACGGTTGGAGAGCTAGCACCTGGTGGTTAAGCATACCTCTTACGTATGATCAAGAGAACTAAAAGCAGCTCCATGCTACTATCTAGATC 500  
TCTAAGTCCCAACTCTCGATCTGCACCAACCAATCTCTAATGGAGAAATGCAATCTCTCTTCTGATTTTGGTCCAGGTACCATGATAGATCTAG 500

111: R L T V G E L A R G W V K D Y P L T Y D E E K L K A A P M Y Y L D P 144

501 CAGCAACTGGCTGGCAAAACCTTGGGAACAAATGCTACTCTCTCTCATCAGGAGCTATGGTAACTGGCTGCTATCAAGATGGTTTAACTGGTACTA 600  
GTGTTGACCGGACCTTTGGAACTCTTTTACCATGATGGAGGCAAGTAGTCTCTGATACCAATTGACCGACCATAGTCTTACCAAAATTGAACCATGAT 600

145: A T G W O N L G N K W Y Y L R S S G A M V T G M Y Q D G L T W Y Y 177

601 CCTAAATGCAGGTAAATGGAGACATGAAGACAGCTTGGTTCCAAGTCAATGGTAACTGGTACTATGCCATGATTCAGGTGCTTACCTGTTAATACCA 700  
GGATTTACGTCCATTACTCTGTACTTCTTCCAACCAAGCTTCACTTACCATGACCATGATACGATACTAAGTCCAGAAATCGACAAATTATGGTGT 700

178: L N A G N G D M X T G W F O V N G M Y Y A Y D S G A L A V M T T 210

721 GTAGGTGGTACTACTTAACTATAATGGTGAATGGTTAAGTAATGAAGGCTAATGTGAACTGTGATGGATACTTAACTTTGTATAATAGGTGGATAA 800  
CATCCCAATGATGAATTTGATATTACCACTTACCAATTCATTACTTCCGATTAAACATTTGACACTACCTATGAATTGAACATATTATCCACCTATT 800

211: V G G Y Y L N Y N G E M V K . 225



9921561

- 13 -

(SEQ ID NO: 32) : TTTATGATATTTATATTAAGAAAGCCATTATTCACAGTTCAGTCCGGATGATACCGAGCTGTTCTTAGCAGATAAGTTTCTCAATATTAAGTCCAAAA 100  
(SEQ ID NO: 33) : AAAATACCTATAAATATAATTTCTTTGGTAAATAGTGGTCAAGTCAGGCTTACTATGGCTCGACAGAAATCGTCTATTCAAGAGTTATAATGAAGTTT 100  
(SEQ ID NO: 31) : M D I Y I K K A I I N O F S P D D T E L F L A D K F L N I T P K 11

101 ATGGAAGAATACCTACGTAAGAAAAATTGAACATGTGTATTGAGATGAAGCAAGACTGGGATTTCGAAGAAGAAAAATCCCTTTCAATCATATTACAG 200  
TAGCTTTCTTACGATGCAATTTTAACTTGTACACATAAGTCTACTTGGTCTGACCTTAAAGCTTCTCTTTTGAAGGAAGTTAGTATAATGTG 200  
33 I E E Y L R K K I E N V Y S D E A R T G I F E E E M P F F N H I T D 66

201 AGGATTTCTTGGAGACATCAGTAACTCTGTAATCTCTGAAAACAGGATTTAGCATTCTCAAAATCTCAAGACCAATGACTTGATTTCTTCAATT 300  
TGCTAAACAACTCTGTAGTCAATGCAACCGATTAGAGAGCTTTCTCTCAAAATCTTAAAGACTTTTAGAGTTCTGCTTACTGAACATAAAACAAAGTTAA 300  
67 D L L E T S V T L A N L W K E E F S I S E N L K T H D L I F V Q F 99

301 TTTAAAGAAGGTGACAGATTTCCTTTCTTGGCAATTCCTCTCGGAGAGCTTCAACCCACTCGGAGGAGAAGTTGATAATCAATCAAGCTGACT 400  
AAGATTCTTCCACATCTTTAAAGCGAAGAACCTTAAACGGAGCCCTCTGCAACTGGGTGAGCTCTCTCTTCAACTATTAGCTTAGTTTGGAGTGA 400  
100 S K E O V E N F A F L R I A L E T L T N L G G E V D M P I K L T 132

401 CAGATAAAGCTGCTGATTTGGAACGGGTCTCAGGAGGCTTGGTGGTCAATCTTCAGAGTCCCAAGTATCACTGATCAAAAAAGCAATCAAGTACA 500  
CTCTTATGGAAGGACTTAAACCTTGGCAGGACTCTCTCGGACCAACAGTATAGAGTCTCAGCTTCAATGCGAGTAACTTTTCTTGAATTCATGT 500  
133 O N N L P G F G T C A D E A L V V N L O S R K Y N L I E R R I K Y N 166

501 ACGGACTTTTCAACTATTTTTCAGATAATCTTTCTGCTGCTCTTAAAGTTTCTCTTAAAAATCTATCAAGGAATCGAAAAAAGCAGCCAGAG 600  
TCCCTTCAAAAAATTCATATAAAGTCTATTAAGAGACGACAGGAGGATTTCAAGAGGATTTTATAGATAGTTCTTCTGAGCTTTTCTTGGGTCTC 600  
167 C T F L N Y F S C H L L A V A P R I S P K K S I K E L E K T A Q R 199

601 AATGCTGAATCTTTTAAACAGATGATTTTCAATTTCAATCCAGGTCAATCAGCTATTTTCAACAACTAGAGAGAAAGCAATGAATTTGACCTGAG 700  
TTAAGCACTTAGAAAAATTTGCTCTACTAAAAAGTTAAAGTTAGCTTCCAGTTTACTCGATAAAAGTTGTTGGATCTTCTTCTGCTTACTTAACAGTGGACTC 700  
200 : A E S F N T D D F Q F O S R V K S A I F N N L E E S N E L S P E 232

701 AAATGCTGAATGACCTTTTGAACAAATCTGACGGTCTTTGAGCTTTATGACCAAGTCAAGAGAGCTTACCAGAACTGTTCAATTTGATGAAA 800  
TTTAAACCAATTAAGAAAAATGTTTGTAGACTCCGAGCAAACTCAAAATACTGTTTCACTCTTTTGGCATGCTCTTGGCAAGTTAAACTACTTT 800  
233 K L A N D L F D N K L T A R L S F I D O V R E A V P E P V O F D E I 266

801 TCGATGCACTCCCAATTAAGAAAAATTAAGAAACCAAAATCTCTTATCAAAATGCAATTCAGCTCATGCTTCCCAATAAGCTTATCAAGAGCGGA 900  
AAGTACGCTCAGCGCTTAATTTCTTAAATTTGCTTTTTCAGAGCAATAGTTTACCTTAACTCAGTAGCAAGGTTATTCCAGATAGTTCTGCGGCT 900  
267 C A S P O L K K F E N O K L S L S N G I E L I V P N W V Y O D A E 299

901 GCTGCTGACTTATCCAAACGAAAAATGGAACCTACTCTATCTTAAATCAAAATATGAGGATATCCAAAGTAAATAATGTTTAAACCAATTCGAAGAG 1000  
CAGACAACTCAATAGCTTTGCTTTTACCTTGCATGAGATAGAAATAGTTTATAGCTCTATAGCTTTCATTTATTACAAATTTGCTTAAAGCTTCTC 1000  
300 S V E F : O N E N G T Y S I L I K N I E D I O S K . 323

1001 TGTGCTACTAGCAGTCTCTCTTTCTGCTCTATAAAGCTTACCGGTTTATCAAGATGTCAAAAGTTCATGACCTATCAACCAATGCTGCGAGAAAT 1100  
ACGAACATGATCTCAGAGGAAAAACGACCGATATTTCAATGCGCAAGTAGTTCTACAGTTGTTTCACTACTGATAGTTGGGTACCACTCTCTTTA 1100

90P1380

- 14 -

(SEQ ID NO: 35)	1	AAATGCTATAAATACAGAAAAATCTTGTGGAGTTCCATTATGGCAATATTTTCATGATTTTTCGATTGTTTGTGTGCTCTTATGCTGATAGCT	100
(SEQ ID NO: 36)		TTTACACGATATTATGATCTTTTATGAACACCTCCAGGTAAATACCTTTATAAAAGTACTAAAAAGACTAACAAACACAGGATAAACCACTATGAC	
(SEQ ID NO: 34)	1		
		MA : F P M I F L I V C V L L L V I V	19
	101	ACACTGACTACACTTTATGTGGTTCTGTCAGCAGTCGGTGGCATTTATGAAACCTTTGGGAAATACCAAAGGTTGCTAATAGCCGTATTCATATTGCT	200
		TGTGACTCATGTCAAATACACCAAGCAGTCTGAGCCACCGCTAATACTTGGAAACCTTTATGTTTTCACAGATTATGCCATAAGTATAAGGCA	
	20	T L S T V Y V V R Q O S V A I I E R F G A K C Y O K V A N S G I N I R L	53
	201	TGCTTTTGGGATTGACTGGATTCAGCACGGATTCACTTGCCTTGTTCGAAATCATATTGTGTTGAGACTAAGACCAAGGCAATGTCTGTTTAT	300
		ACCGAAACCTTAAGTACGCTAAGCTGTGCTTAAGTCAACGGCAACACCTTTCAGTATAACACCAACTCTGATTGTGTTGCTGTTACACAAGCAATA	
	54	P F G I D S I A A R I Q L R L L O S D I V V E T K T K D N V F V H	86
	301	GATGAATGTAGCACTCAGTACCTGTCTCAACGAGCAGAGCGTGACAGATGCTTACTATAAATCATACCTCCAGAACTCAGATTAATCTTATATGAA	400
		CTACTTACATCGCTCAGTCAATGGCAGCTGCTCTCGCACTGTCTACGAATGATATTGAGTATGCAAGGCTCTAGAGTCTAATTTAGAATATAGCTT	
	87	M N V A T O Y R V N E Q S V T D A Y Y K L I R P E S O I K S Y I E	119
	401	GATGCTTTGCTCTCTGTTCCAAAATTAACCTTGGATGAATGTTTTCAGAAAAAGATGAGATTGCTTTGAGGTTCAACACCAAGTAGCAGAGAAA	500
		CTACGAGAAGCCAGAACAGCAAGGTTTAAATTGGAACTACTTTAAGCAACTCTTTTCTACTCTAAGCGGAACTCCAGTGTGTTGTTTATGCTCTCTT	
	120	D A L R S S V P K L T L D E L F E K K D E I A L E V O N O V A E E H	153
	501	TGACCACTTACCTACATTATCGTGAAAACTTGAGTTACCAAGTCGAAACCGATACGCTTAAAGCAATCTATGAATGAAATCAATGCGGGCAAG	600
		ACTGTTGAATGCGGATGTAATAGCACTTTGGAACTAATGCTTCAGCTTGGCTCATGCTCTTCAATTCGTTAGATACTTACTTTAGTTACGCGCGCTTGC	
	154	T T Y G Y I I V K T L I T R V E P D A E V K Q S M N E I N A A O R	186
	601	TAAAGCGGTCGACACAGAAATGGCGGAAGCTGACAGATTAAATTTGCTCACTGACCTGAAGCGCAAGCAGAAAAAGACCGGCTTATGTTGTGGGG	700
		ATTGCGCCAGCGTGTGTTCTTAAACCGCTTCAGCTGCTGAATTTTAAAGTCAGCTCGACTTGGCTTCTGTTTTCGCGGGAAGTACCAACCCCC	
	187	K R V A A Q E L A E A D K I K I V T A A E A E A E K D R L H G V C	219
	701	ATTGCCCAACAACTAAGCCGATTGTGATGATTTGGCAGAGTCTATCACCGAAGCTCAAGGAAGCAATGTTGGTCATGACAGAAGAAATCATGTTCTA	800
		TAAAGCGGTTGTTGATTTGCTGAACACTTCTTAAACCGTCTCAGATAGTGGCTTGAAGTCTCTTGGTTACAAACGATAGTCTCTCTTGTTTAGTACAGAT	
	220	I A O C R K A I V D G L A E S I T E L K E A N V O M T E E O I M S I	253
	801	TGCTTTGACCAACCACTATTGGATACCTTGAATACCTTGGCTTAAAGGAATCAAAACATCTTTTACCAATACTCCAAATGTTGTGGATGATAT	900
		AGGAGAACTGTTGCTGATAAACCTATGCAACTTATGGAACCGCAGATTTCTTTAGTTTGGTAGAAAAATGGTTTATGAGGTTTACCAACCTACTATA	
	254	L L T N O Y L D T L N T F A S K G M O T I F L P N T P N C V D D I	286
	901	CGGTACACAAATCTTTCAGCGCTTGGCTGAGAAGAAATATAGACTAATACTCTTGGAAATCTCTTCAAACCTACGTCAGCGTCTGTTGCGGTATA	1000
		GCGATGTTGTAGAACAGTCCGGAAGCGGCACTCTCTTTATTTATCTGATTAGAGAAGCTTTAGAGAAGTTTGTATGAGTCAGTCGACGACAGAACCGCATAT	
	287	R T O I L S A L R A E K K	300

seq1713

- 15 -

(SEQ ID NO: 38) : CTTGTATATGTTGATAAAATAGGCTTTTATTTCGAAACCTTTCTTTGTTTCAAAATGCTAAAAAATGCTACATAGAGCAAGCTTACTATTA 100  
(SEQ ID NO: 39) GCACTATACCACTATTTTATCCCAAAATTAACCTTTTCGAAAGCAACAAAGTTTAAAGATTTTACCACTGTTATTTCTTCCAAATGATAAT

101 TCTGAATCAGCAGATTTGACACAAAGATTCATTTTGAATCAATAGGCTTTATGAAAACCTGAAGGCTTCTCTAGTAAGACCTGATTTTATTCGG 200  
AGACTTAGTCTCTAAACCTCTCTTTCTAAGTAAACCTTTAGTTATCCGAAATTAACCTTTGCACTTCCCAACAGATCTTTCTGCACTAAAAATACCC

(SEQ ID NO: 37) 1 L R S I G F I E K L R G L S S K E L I L L G 32

201 AATTATCCTAAGTATCTTTTACCTTTTATCTTTCTAGTTTATCTCTGTTTATATATTATCACTTTGATTTTACAGGACACATGAAAAGTATTTCTT 300  
TTAATAGGATTCATAGAAAAATGCGAAAAATAGAAAAACATCAACATCAGACAAATATATAATAGTCAAACTAAAAATGCTCTCTGATCTTTTCATAGAA

23 I I L E I F L P F Y L F V V V L C L Y I I S L I F T G D M R S I L 55

301 CAGAAAAATCGGGGAGCATCCGATGCTGCTTTCTTACCTATAGTACTGTTATATCCATCTTTCACAAAAATGCAATGCTCTTCTGCTTCACTAG 400  
GTCTTTTACCCCTCTGAGGCTACGAGCAAGAAAAAGATCATATCATGACAAATATAGTTAGAAAGCTGTTTAACTACCCAGAACACCGAAGTCACT

56 O K M G E R P M L L L F L S Y S T V I S I L A Q M W M G L V A S V G 89

401 GAATGTTTCTATTACTATTTTCTTTTGCATATCAGTGGATTTTATCCATAAATCTTTGATTTGATTTTTCAGTCTGCTCTTTGCTAGTGTCTT 500  
CTTCAAGATTAATGATAAAAGAAAAAGTGTATAGTCACTAAAAATAGGCTATTTAAGAAAGCTAACTAAAAAGTCAAGCAGAACAAACCTACAGAA

90 M F L F T I F F L N Y O S I L S N R F P R L I L Q F V L F G S V L 122

501 GTCACTGCTTTCCCACTTTAGAACATTTCAAAATGTAAGAAATTAACATGCTTTTCTTCCCAATATGCAAGTGTGCTATGACAAACCCGCA 600  
CACTGCAAGCAAAACCGTCAAACTCTGTAAAGCTTTAACACTTCTTAAATGTAAGCAAAAGAAAGTGGCTTATACCTCCACACCTGATTTGCCCCCT

123 S A A P A S L E N F C I V K K F N Y A F L S P M M Q V M M O M R A 155

601 GAAGTGAATTTCTTAATGCTAATTTATGCAATTTTCTGTTTCTGTTATGATGCTTTCTATCTGTTTACAAAGACCAAGTGAATGCTGGA 700  
CTTCACTGGAAGAAATTAAGATTAATAAATGCTTAATAACCAAGCAAGACATAATCTAAGCAAGATAGACAAATGTTGCTGCTTCACTTAACCAACT

156 E V T F F M P N Y Y G I I C C F C I M I A F Y L F T T T K L M W L K 189

701 AAGTATCTCTGCTGATTCAGGCTTTTAACTCTCTTCTGTTTGAATTTACTCAAAATCCAACTGCTTTCTGCTATTTATGCTGCAAGCAATTTCTA 800  
TATCAAGACACACTAAGCTCCGAAACAAATAGAGAAACCAAACTCAAAAGCTTTAGCTTGACGGAAGGACGATTAAGCAAGCTCTTAAATAGAT

190 V F C V I A G F V N L F G L M F T O M R T A F P A I I A G A I I Y 222

801 TCTTTTACGACTATTAAAACTGGAAGCTTTTCTTACTATTGCGCTTCCGCAATGCTTTCAGTCTTCTCTTTTCTAGTGAATTTGCGAGTTGCA 900  
ACAGAAATGCTGATAATTTTTCAGCTTCCGAAACCGAATCATACCCCAAGAGCGCTAACCAAACTCAAGAGCAAGAAATCACTAAACCTCAAGCT

223 L F T T : K M W K A F M L S I G V F A I G L S F L F S S D L G V R 255

901 ATGGTACTTTAGACTCTCTATGCAAGAACCAATTTCTATCTGGATGCTGGATGCTTCTTTAAGCAAAATCTTTTGGGCTGAAGGGCCATTTGA 1000  
TACCCATGAATCTCAGAGATACCTTTCTGCTAAGATAGACCTTACGACCTTACCGGAACAAATCTGTTTAGGAAAAACCCCACTTCCCGTAACCT

256 M G T L D S S N E E R I S I N D A G H A L P R O M P F M G E G P L T 289

1001 CTTATATGCACTCTTATCTCTGGATACATCTCTTATCATGAACATGCCACAGTCTTTATATTGATACGATTTCTGAGTACGGAATTTGCGGTACCAT 1100  
GGATATACGTTGAGAAATAGGAGCTATGTAAGGAAATAGTACTTGTACGGGTGTCAGAAATATAACTATGCTAAGACTCAATGCTTTAACACCCATGCTA

290 Y M H S Y P R I H A P Y H E N A N S L Y I D T I L S Y G I V G T I 122

1101 TTTATTAGTTTCT 1200  
AAATAATCAAAACAGAGACAAAGGACAAAGCAAGCAACT

123 L L V L S S V A P V R L M M D N S Q E S G K R P I I G L Y L S P L 1355

1201 ACAGTGTCTCTGTCACGGAATTTTCACTTGGCT 1299  
TGTCAACCAACCACTGCTCTTAAAACTGAACCGAGAGAGACTTAAGTCACTCCGAAATAAAGAACGATCAATACACGCTCAATAGGTAACCGAAAT

156 T V V A V M G I F D L A L F M I O S G F I F L L V M C S I P L A L 1388

gcp223

- 16 -

(SEQ ID NO: 41) : AAGGAGTGAAACATCTGGCTGGTACTTCAATTGATGAAAGTATGGTGAAGAAATTCGTGTAAACAGTTTGGCAACGGGTGTCTGCAAGACCGGCTAGA 100  
(SEQ ID NO: 42) : TTTCTCACTTTGTAGACCGAGCCATGAAGTTAACTACTTTCAACGCACTACTTTAAGCACATTTGCAACAGCGTTGGCCACAGCACTTCTGGCCATCT

101 AAGGTTGTGGCTCCACAGCTAGATCTGCTACTAATACCTGAGACAGTGAACCCAGCTCATTACATGGCTTTGATGCTCAATTCATATGCCAGAA 200  
TTTCCACACCGAGGTGTCGATCTAGACGATCATTCATGCGACTGTGCTACTTTGTTGAGTAAGTGTACCGAACTAGCAGTAAACTATACCGTCTT

201 AAGTTGAATTGCCAAACAAAAATCCAGCTGTTTGGAAACCACTCAGGCACTGCTTTGTTGATTTGGGATCTTCCGCTTAATGCAATGTTCTGTACAA 300  
TGTCACTTAACGGTTTGTTTTAGGTGCAGCAAACTTGGTTGAGTCCGTAGAACGAAACCACTAACCTTAGAAGCGGCACTTAGCTAACAGCACTGT

301 CAGATTCACTGCTTTTCCAGTCCAGCGCTTTGAAGCCCCAATTCACAGATCAGATGAATTCGATACACTTCCATTTTCAAAAATGTTAAGTAAA 400  
GTCTAAGTCAGCAAGACGTCAGCTCCGAACTTCCGGTTAAAGTGTCTACTTCTACTTAACCTATGTGGAGTTAAAGATTTTATAGCAATTCATTT

(SEQ ID NO: 40) : M 1

401 TGAATGTAAAGAAAAATACAGAACTGTTTTCGAGAACTGCGAGCGCTAGTCTGAGTCTCTCATGGAGAGAGTGTGCTGCTCTGCTCATTTGCACTTAT 500  
ACTTACATTTCTTTTATGTCTTCAACAAAAGCTGTTCAACGTTCTGGATCAGACTCAGCAGTAGCTCTCTCACCAGCCAGAGACAGTAAAGTCAATA

2 N V K E N T E L V F R E V A E A S L S A N R E S G S V S V I A V I 14

501 CAAGTATGTAGATGTACCCACAGCGGAGCTTCTGCTCCGCTAGGTGTTCACTATATCGGTGAAGATCGTGTAGATAAGTTTCTGAAAAATATCAAGCT 600  
CTTCATACATCTACATGGCTGTGCTTCCGAAACGAGGCGATCCCAAGTAGTATAGCCACTTTTAGCACATCTATTCAAGAGCTTTTATAGCTTCCA

35 K Y V D V P T A E A L L P L C V N H I G E N R V D K F L E K Y E A 67

601 TTAAGATCCAGATGTGACTTGGCATTGCTGCTGCTTCCAAAGACGTAAGGTGAAGATGTCAATACGTTGATTATTTCCATGCAATGGACT 700  
AATTTCTAGCTGTACACTGAACCGTAACTAACGATGGAACGTTTCTGCACTTCTTACAGTAAGTTATGCAACTAATAAGGTACGTAACCTGA

66 L K D R D V T W H L : G T L O R R K V K D V I O Y V D Y P H A L D S 101

701 CAGTAAAGCTAGCAGCGGAAATTCAAAAAGAGTACCGAGTCAAGTGTCTCTCAAGTAAATATTTCTAAGAGAAAGCAACACCGGTTTCT 800  
GTCAATTCGATCGTGGCTTAAAGTTTTCTTCACTGGCTCAGTAGTTCAAAAGGAAGTTCATTTATAAGATTTCTTCTTCTGTTGTCGCAAAAG

100 V K L A G E : G K R S D R V I R C F L O V N I S K E E S K H G F S 114

801 GAGAGAGGAAGTGTGGAATCTTCCAGACTTAGCCAGACTAGATAAGATTCAATATGTTGCTTTAATCAGCATGGCACCTTTGAGGCTAGCAGTGA 900  
CTGTCTCTGAGCACTTAGAACGGTCTCAATCGTCTGATCTATTCACTTATACAGCAAAATTAAGTCTACCGTGGAAAGCTCCGATCGTCACTC

135 P E E L L E : L P E L A R L D K I E Y V G L N T H A P P E A S S E 167

901 CAGTGAAGAGATTTTCAAGCGGCCCCAGATTACAAAGAGAAATTCAGAGAGAAACAAATTCAAATATGCTTTAGAGCACTGCGCGGCTTAC 999  
GTCACTTTCTTAAAGTTCCGCGGCTTCAAAATTTCTCTTTAAGTTCTCTTTGTTTAAAGTTTATACGGAATCTGTTGACCGCGGCAATG

140 C L K E : F R A C D L O R E I O E K O : P N M P L E N T G R Y 200

- 17 -

[illegible]

99p273

- 18 -

(SEQ ID NO: 47) 1 CAATGTGTCCTGAACTTTTACAAAACATCTTCTGAAAAAGAGTTGAACTCAAGACCAATTTGTTCAAAATAGGATGCTTGTGCTGATGATG 100  
(SEQ ID NO: 48) 1 GTTACACAAAGGCTTGAATAATCTTTTGTAGAGGAATCTTTCTCAAGCTTGTGAGTTTCTGTTTAAAGCACTTTATCTCTACCAACCAACTACTAC 100  
(SEQ ID NO: 46) 1 M H 2

201 CACAGGATTACACAAGCTTGGAAAAGGTTGGAGCTGCTCTACCTACAGAGACTGTTTATGCTTTTCCAAAGGCTTATGATGAAAAGCAGTTG 200  
CTGCTTAATCTGCTCTCAACTTTTCCCACTTCCAGAGCAAGATGGATGCTCTGCAAAATACAGAAAAAGGTTCCGGAACTACTTTTCTGCAAC 200

3 D R I R O E L E X G C A V V L P T T E T V Y G L F S K A L D E R A V D 36

201 ACCATGTTTACCAACTCAAACGTCCTCTAGAGATAAGGCACTCAATCTCAATATGCTTTTCCAGGACCTCTTGCCTTTTCAAAGAACTCAGCCAGC 300  
TGATACAAATGCTTGAAGTTTGCAGCAGGATCTCTATTCCCTGAGTTAGAGTTATAGCGGAGAAAGCTCTGTAGAAAGTGAAGAAATTTCTTACTGCTG 300

37 H V Y Q L K R R P R D R A L N L N I A S P E D I L M F S K N O P A 69

301 TTATCTACAAAACCTTGTAGAGACCTTTTCCCAAGTCTCTGACCATTTCTGGAAGCCAATGACCGAGTTCTCTATTGCGTAATTTCTGCACTTCTCA 400  
AATAGATGTTTTTCAACATCTCTGAAAAAGGTTCCAGGGAAGTGTAAATAGAGCTTCCGTTACTGCTCAAGGGAATACCACTTTAAGACTGGAAGCT 400

70 Y L O R L V E T F L P G P L T I I L E A N D R V P Y W V N S D L A 102

401 ACTATTGCTTTCCGATGCCAGTCACCTTACCACTGGATTTAATTGAGAGACAGGTCCTTCTGATTGCGGCTCTGCAATATCTCAGGTCAGGCA 500  
TGATAACTTAAGGCTACGGGTCAGTGGGATAGTGTGACCTAAATTAAGCTCTCTGTCAGGGAAGTAAACCGGAGAGAGCTTATAGAGTCCAGTCTGTT 500

103 T I C P R M P S N P : T L D L I R E T G P L I G P S A N I S G O A S 136

501 GTGCTGAACCTTTCAACAAATTTCTGAAGGATTTTGAACCAAGAGGTTCTGGGCTCTGGAAGACGATGCTTTCTAACTGACAGGATTTCAACTATTGGA 600  
CACCACATGGAACTTTGTTAAGACTTCTTAAAGCTGCTTCCAAAGACGAGACTCTGCTACGAAAGAGTTGACCTGCTTAACTGATTAACACTT 600

137 G V T F E O I L K D F D O E V L G L E D D A F L T G O D S T I V D 169

601 TTTGCTGAGACAGGTTGAAAATCTTACCAAGGCGCAATTAACGAGAGATATTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 700  
AAACAGACTCTCTTCCACTTTTGAATGCTTCCGCTTAATTTCTCTTATAGAAAGAGGCAAGGCTCTCTAAGAAAACTCTCTGCAACTTTTAC 700

170 L S G D K V X : L P K A Q L N E K I F L L G C Q R F L L R R L E N 202

701 CTAGAGATTTGCAAGAAACAGATGTGAAGCGATATGTGACATCAACCAAGGCTTTGCTTATCTTTAGTCCAGAGAAACGGCTAGCCAACTAG 800  
GATTTCTTAAAGCTTCTTCTCTACACTTTGCTATACACTGTAGTTGCTTCTCCGAAACCAATATGAAATCAGGCTCTCTTTGCGGATCGGTTGATC 800

203 L R D L O E T D V K A : C C I N O E A L G Y T F S P E E T A S O L A 236

801 CTAGAGTGTCTCAGGATTTCCATCTTTCTTACTTGGTATCAGGATGCACTAATCATCTTTACTTGGATATGCTCCAGCTGCAAGTTTACGAATCACT 900  
GATCTGACAGAGTCTTAAAGGATGTAAGGATGAACCGATCTCTTACTGATAGTACAGAAATCAACCTATACAGGTCGCACTTCAATGCTTATGTA 900

237 R L S O D S N H F L L G Y E D A A N N V L L G Y V N A E V Y E S L 269

901 CTATTCCAAAGCAGGATTTAATATCTTACTTTAGCCTTTAGCAGCTTTCACTCAAGCGCAAGGTCAGGATTCGTAAGGTTTACTACAAAGGTTGCAACAGAA 1000  
GATAAGGTTTCTCTTAAATATAGAAATCGAAATCTGCAAGAGGAGTTTCCGTTCCAGTCCATAGCCATTTTCAAATGATTTTCCCAAGCTTTGTTCTT 1000

270 Y S K A G F N I L A L A V S P O A G G G I G K S L L O G L E O E 302

1001 GCAAAAGATGCTGTTATGCTTTATCCCTTAAATCTGCAATCATCTCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1100  
CGTTTTCTACACCAATACCCAAATAGGCAATTTAAGACGCTTACTAGCAGACCCAGGAGTACGTAAGAAATCTTTTCAACCGATATGAACACTATTTT 1100

303 A R R C G Y G F I R L N S A M N B L G A N A P Y E K V G Y T C D K N 336

1101 TGCAGAAAGGCTTTATTCGATCTTTAGTTTGAATTTCTTATTGTAATCAAACTAATGAGTACTCACAATAAAGGAGAGAGCTTATGATTTTTC 1200  
ACGCTTTTCCAAATAGGCTAGAAAACTAAACTAAAGAAATAACATTTTATTTGATTACCTGATCAGTGTGTTATTTCTCTCTGATAGTAAAAAC 1200

337 O R R F I R I P . 345

999386

- 19 -

(SEQ ID NO: 50) : AAGATAATAGAAAATAGAAATGTAAACCAATGAGAGAAAAATGGCATTGGAGATAATGGAAATCTAAAAAACTATCTTTGAGAAAAATACCTTTTAT 100  
(SEQ ID NO: 51) : TCTATTATCTTTTATCTTACATTTGCTTACTCTCTTTTACCTGTAACCTCTATTACCTTTAGCATTTTGTATACAAACTCTTTATTGGAACAAATA

101 CCGTATTATCATCTAGCAAGCTTTATGGGAATTTTGCACCTGCAATTTGCTGCTTCACTAATCTATAAAATTCATTCAAGAAAAATTTAGTGACTG 200  
GCACTAATAGTACGATCATCTTCAATAACCTTTAAAAACCTTGACGTTAACCCAGCAAGTCATTAGATATTTTAACTAGTCTTTTAAATCACTGAC

201 CGATTCCCGAGCTTTTAAAGTCAGAGAAATTAATGAGTATGTTTTAGATACAGCTAAGATTAAAGTCAAGCTGCTAATGCTGCTATGCTATGG 300  
CTAAAGGGTGGGAAAAAATTTCACTCTTTTATTACTCATACAAAAATCTATGTCATTTCTAATTTCCAGTTCCGACCATACCACTGCTACCATACC

(SEQ ID NO: 49) : M P L D T A K I R V K A G N G O D G N V 20

301 TTGCTTTCTCTGCTGAAAAATATGCTCTAATGGAGCCCTTGGGCTGCTGATGCTGCTGCTGAGGCAATGCTGCTTCTGCTGAGCAAGGACTACG 400  
AAGGGAAGCAGCACTTTTATACAGGATTAACCTCCGGAAACCCACCACTACCAACAGCACTCTGCTTACAGCAGAGCACTCTGCTTCTGATGC

21 A F R R E K Y V P M G C P M G C D G G R G G N V V P V V D E G L R 53

401 TACCTTCATGGAATTTCCGCTACAATCTCTCAAGCT 500  
ATGGAATCACTAAAGGCGATGTTAGCAATAAGTTCCGCTAAGACCACTTTTCCCTACTGCTTTCCCTAGCTTACAGCAGCACTCTCTGGAATCT

51 T L M D F R Y N R N F K A D S G E R G N T K O H N G R G A E D L R 86

501 GTTCCAGTACCACAAGCTACCACTGCTGCTGATGCT 600  
CAAGCTCATGCT

87 V R V P Q C T T V R D A E T C R V L T D L I E N G O E P I V A N G G 120

601 GTGCT 700  
GAGCACTGCT

121 R G G R G N I R F A T P R N P A P E I S E N G E P C O E R E L Q L 153

701 GGAATAAAATCTTGGCAGATCTGCTTATAGGATTCCTATCTGAGGGAAGTCAACACTTTAAGTGTATTACCTCAGCTAAGCTTAAATTTGCT 800  
CTTGATTTTATGAACCTGTACAGCAAAATCATCTAAGGCTAGACATCTCTCAGTTGTGAAAAATTCACAATAAGGAGTCGATTCGATTTTAAACA

151 E L K I L A D V G L V G F P S V G K S T L L S V I T S A K P K I G 186

801 CCTACCACTTTACCACTATTGTACCAATTTAGCTATGCTTGGCAGCAATCAAGTGAATCTTTGCACTAGCCACTTCCAGCTTTGATTAAGGGG 900  
CGGCTGCTGAAATGCTGATAACATGCTTTAAATCCATACCAAGCTGGGTTAGTCCACTTAGCAAACTCATCGCTGAAAGCTTCAAACTAAGCTTCCCC

187 A Y N F T T I V P N L C M V R T O S G E S F A V A D L P G L I E G A 220

901 CTAGTCAAGCTGCTGCTTGGGAACTCAGTCTCTGCT 1000  
GATCAGTTCCACAACCAACCTTTCAGTCAAGGAGGAGTGTAGTCCGATGCTGCAATAGGAAGTGTAGTAACTATACAGTGGATGCTTCCGGCACT

221 S C G V C L C T C F L R N I E R T R V I L N I I D N S A S E G R D 253

1001 TCCATATGAGGATTACCTAGCTATCAATAAGAGCTGAGTCTTACATCTTCCCTCATGAGGCTCCACAGATTATGTAATAATAGATGGACATG 1100  
AGGTATCTCTAATGGATGATAGTTATTTCCAGCTCAGAAATGTTAGAAAGGAGTACCTCCAGGCTGCTAATAACATTGATTATTCTACCTGTAC

251 P Y E D Y L A I N K E L E S Y N L R L M E R P Q I I V T N K M D H 286

1101 CCTCAGCTCAGGAAATCTTGAAGAAATTAAGAAAAATTTGGCTGAAATTTATGATGAATTTGAAGAGTTACCACTATCTTCCCAATTTCTGGATTGA 1200  
GAGCTCTCAGTCTTTAGAACTCTTAAATCTTTTAAACCACTTTTAAATCTAATTAACTTCTCAATGCTGATAGAGGGTTAAAGACCTAATCT

287 P E S O E M L E E F R K K L A E N Y D E F E E L P A I F P I S O L Y 320

1201 CCAAGCAAGCTCTGCAACACTTTAGATGCTACAGCTGAAATTTAGCAAGAGACCAAGAAATTTTCTCTACCAAGAGTCCGATATGCAAGAGAAAT 1300  
GCTTCTTCCAGAGCTTTGCAAAATCTAGATGCTCACTTAAGAAATCTGTTCTGCTGCTTAAAGAGAGATGCTGCTCAGGCTATACCTTCTCTTCA

321 F O C L A T L L D A T A E L L D K T P E F L L Y D E E D N E E E V 353





989111

- 21 -

(SEQ ID NO: 53) 1 TCGAATGCTCTTAAGAAAAAATTCAGAAATCAAGAAAAACAGTAAGACAAGTCTCTCTATGAAATATTAGAAATCAAGAAAAAAGCATATTAT 100  
(SEQ ID NO: 54) ACCTTACGGCAATCTCTTTGTTAACTTTAGTTCTCTCTCATCTCTGTTCAAGAAAAACAGAAATATTAAATCTTTACTCTCTCTCTATATA 100  
(SEQ ID NO: 52) 1 M 1

101 GCGTAAGAAAGAGTAGAACCAAAACCAATTGACCTTGGTGAATATAAATTTGCTTTCCATGACCATGTAGAGCTTGTCTTATCGACAGCAAAAGGACTT 200  
CGACTTCTCTCTCATCTCTGTTTGGTTAACTGGAAACCACTTATATTAAACCAAAAGCTACTCTACATCTCGACAGAAATAGCTGTCTCTCTCTAG 200  
2 A E E R V E P E P I D L G E Y K F G F R D D V E P V L S T G R G L 34

201 AAGCAAGGTGTTATTGTTGAATATCTGCTCTAAAGGTCAGCTTGAAGTGTGGAGTTTGGAGTTTGAAGTCTTATGAACCTTCAAAAAAATGCCCA 300  
TTCTTCCACATAAGCACTTAAATAGACGACGATTCCCACTCGGACTCAGCTACAGCTCAAGGCAAACTTCAGAAATCTTGGAGTCTTTTACAGGTT 300  
35 M E G V I R E L S A A K G E P E N H L E P R L E S Y E T P K E N P M 40

301 TCGAACTTGGGAGCAGACTTGTGAGAGATTGACTTTGATGACTTAATCTACTACCAAAAACCTCTGACAAACAGCCCTTCTGGGATGATGATCC 400  
ACCTTGAACCCCTCTCTGAAACAGTCTCTAACTGAAACTACTGAATAGATGATGCTTTTGGTGAAGTCTTGGTGGGCAAGAACCTTACTACATGG 400  
69 Q T M G A D L S E I D F D D L I Y Y Q K P S D K P A R S W D D V P 101

401 TGAAGATTAAGAAACCTTTGAACGTATCGGATTCCAGAGCTCAAGTCTCTTATTAGCAAGGCTTCTGCGCACTAGAGTCAAGAGTGGTTTAC 500  
ACTTTCTAATTTCTTTGGAAGCTTGCATAGCCCTAAGGCTCTTCACTTCCAGCAATAAATCTGCTCCGAAAGACGGCTCATGCTCAGCTTCCAGAAATG 500  
102 E K I K E T F E R I G I P E A E R A Y L A G A S A O Y E S E V V Y 114

501 CACACATCAAGCAAGAGTTCCAAAATTAAGTATTATCTTTACAGATACAGATTCCGCACTCAAGCAATACCCAGACTTATTTAAACATATCTTCCCA 600  
GTCTGTACTCTCTCTCAAGGTTTTTAATCCATAATAGAAATGCTATGCTCAAGGCGTAGTTCCTTATGGTCTGAATAAATTTGCTCAAGAGCT 600  
135 M M H K E E F Q K L G I I F T D T D S A L K E Y P D L F K Q Y F A K 160

601 ACTGCTACTGCGCAGATACCAAGTTGGCAGCCCTCAACTCAGCAGTATGCTCGGTCGAACCTTTATCTAAGTCCAAAAGGTGTCAAGTAGATAT 700  
TCAACCATGCGGCTGTCTATTGTTCAACCTCGGAGTTGAGTCTCATCTGCAAGCCCACTTGAATAATAGATGCAAGGTTTCCAGAGTCCATCTATA 700  
169 L V P P T D N K L A A L N S A V W S G G T F I Y V P K G V K V D I 201

701 TCCACTTCAAACTTATTTCCGTATCAATAACGAAATATAGGTCAGTTCCAACTACCTTGATTATGCTGATGAGGAGCAAGCTTCCACTACGTAGAA 800  
AGGTGAAGTTGAATAAAGCATAGTTATGCTTTATATCCAGTCAAGCTTGCATGCACTAATAGCAACTACTCCCTCTTCCAGGTGATGCACTCT 800  
202 P L O T Y F R I N M E N : G O F E R T L I I V D E G A S V H Y V E 234

801 CGATGTACAGCACCACATATTCAAGCAATAGCTTACAGCTGCCATGTAGAAATTTTCTTTGGACGGAGCTTATATGCTTATACACTATCCAAA 900  
CTTACATGTCGTGTTGATAAGTTCTTATCCAAATGTCGACGGTAACATCTTTAAAACGAAACCTGCTTCCGAATATACGCAATATGTTGATAGGTTT 900  
235 G C T A P T Y S S N S L H A A I V E I F A L D G A Y M R Y T T I O N 260

901 ACTGCTGATACGCTCTATAAETTCGTAACAAAGCTGCTAAGGCTCAAAAGGATGCCATCTGTCAGTGGATTGATGAAACTTGGGTGCCAAAAGTAC 1000  
TGACGAGACTATTGAGATATTGAACCAATTGTTTCCGACGATTCCGAGTTTCTTACGCTGACAACTCAGCTAAGTACTCTTGAACCCACGCTTTGCTG 1000  
269 M S D N V Y N L V T K R A K A Q K D A T V E W I D G N L G A K T T 301

1001 TATGAATATCCATCTGTTTACCTTGAAGCAAGAGCGGCTGCTACATGCTCTCTATGCTTTGCTAATGCAAGGCAACACCAAGACACCGGTGCT 1100  
ATACTTATAGGTAGACAAATGGAATACCTCTCTCTGCGCCACATGCTACGAGAGATAGCGAAACGATTAAGTCCGCTTGTGCTTCTGTCACCA 1100  
302 M K Y P S V Y L D G E G A R G T N L S I A P A M A G O N O D T G A 114

1101 AAGATGATTCAATGCTCCACATACAGCTCTGCTATTGCTCTAAATCCATGCTAAAGGTGAGGAAAGGTTGACTACCTGACAAAGTCACTTTA 1200  
TTCTACTAAGTTTACGAGGTGATATGCTGAGCAGATAACACAGATTAGGTAGGATTTCCACTCTTTCCAACTGATGCGACCTGTTCAAGTGAAT 1200  
335 K M I N H A P N T S S S I V S K S I A K G G K V D Y R G O V T P M 360

1201 ACAAGAACTTAAGAAATCTGTTTCCCACTTGAATGTGATACCAATTAATGATGATGACTTT 1263  
TCTTCTGAGATTCTTTAGACAAAGGTTGAACTTACACTATGTAATAGTACTACTGAAA 1263  
369 K M E K E S V S M I E C D T I I N D D L 380

gsp3362

- 22 -

(SEQ ID NO: 56) : AGCTGCAATTTATGAGCAAGTATCTATCTTAAAGAACCAAGAGTGTCTTCTAACTCGTTATAATGAAGTTCAAACTCAAAACAGCAACTTTAATCTTA 100  
(SEQ ID NO: 57) : TCGACCTTAATACTCGTTCATAGGATAGAAATTTCTTCTTCTCAAAATAGATTGAGCAATATTACTTCAAGTTTGACTTTGTGTTGAAATTAGAAC  
(SEQ ID NO: 55) : A G I Y K Q V S Y L K E Q R S V Y L T R Y N E V O T E T A T L I L 111

101 GGAGCTATTGTGGGATAGCTAGTTCCTTGTACTCTTTTATCTGTCAATCTTCTATATTTGAGCAATTCGCGGAGATATCTTCATTAAACCAATTT 200  
CTCGATACACCCCTATCGATCAAGGAAACATGAGAAAATAAGACAGTTAGAAAGATATAAGCTCGTTAAGGCGGCTCTATAGAACTAATTGCTTAAAC  
34 G A I V C I A S S L L L F Y S V N L L Y F E Q F R R D I L I E I S 67

201 CAGCTTTACCATTTTTTGAACACATGCTCAGTATATGTTAGTCAATTTGCCAGTTTGTATTTGCTGCTAGTCTCTTTATTTAAGCACTCGAGACTT 300  
GTCCAAATGCTAAAAAACTTTGTGTACGAGTCAATATACCAATCAATTAACGTTCAAAACATAAACCAAGATCAGAGAAATAAAATTCGTCACTCTGAA  
68 G L R F F E T H A Q Y M V S Q F A S P V F G A S L F I L S S R D L 100

301 GGTGATGCGCTTCTCACTTTATAGTCTTTCTAGCTAGTGCAGTTTTCAGCCTTTACGTTCAAGCGCAGAAAGAAATCTGTGTTTCTATGACAATTATG 400  
CCACTAACGCAAGCACTGAAATAATCAGAAAGATCGATCACTGCAAACTCGCAATGCGAGTTCGCTCTTTCTTAGAGCCACAAAGATACTGTTAATAC  
101 V I G L L T L L V F L A S A V L T L Y R Q A Q K E S R V S H T I N 111

401 AAAGCAAAATAGCATGATTGAACATAAGAAATATATCTAAAAAAATTTGGAAGCGTCAGCTATTTTCAGATACCAATCTTTA 481  
TTTCTTTTATCTTACTACTTGTATTTCTATATAGATTTTTTAAACCTTCGGCAGTCGATAAAAGTCTATGCTTAGAAAT  
114 K G K \* 117

gcp3387

- 23 -

(SEQ ID NO: 59) 1 TTTTATCTAGTACAGTATATTTATTCGGCTGTCCCAATATTCATCCATCCAAATGTATTAGAAATGGATCTTATGTTTACTTCAAGATATGACGACTGG 100  
(SEQ ID NO: 60) AAATAGATCATGTCATATAAATAACCGACAGCGTTATAGTTAGGTAGGTTTACATAATCTTACCTAGAAATCAAAATGAAGTTCTATACTGCTGACC  
(SEQ ID NO: 58) : M T T G 4

101 AGTATATTGCTTTCCGTTTCATATATATTCCTTTTATTTGATGAATAACTATTTAATAGGTTGGAGTGTCCATTCGTCGAATCAATTAAAG 200  
TCATATAACGAAAGCCAGGTGTATATATAACCAAGAAAAATAAACTACTTATTGATAAAATTTCCAGCTCAGACGGTAACGACTTTAGTTAATTC  
5 V Y C P P F T Y I L F P F Y L M N Y F N R L S C R I R L K S I K 37

201 CACTTACCACTTTTACTTTCAAATACCACTCTTACTACCGGATTTGGACCGGACTTTATTTTATTGATTTTCTAATTGCATTAGTAATGTT 300  
GTGAAATGCTCAAAATCAAGTTTAATGCTGAGAAATCATGCCCTAAACTCGCCCTGAAATATAAAATAAATACTAAAGATTAACTAAATCATTACCAA  
38 N F T S F S F K L A A L S T G I M T A T L F L L I F L I A F S N O F 71

301 TTAGCTTCTCTTTGGAGATAAAGGAGGTGATTTTAAAGAGAAATTTATGTTATAGTATTGCAACAAATGCTACTTTCTTTATAGGATTTTCTTCTC 400  
AATCGAAGAGAAACCTCTATTTCTTCAACTAAAAATTCCTTAAATACCAATATTCATAACGTTTGTAGCATCAAGAAATATCTTAAAAAAGAG  
72 S F S L E I K E V D F L R E F Y G I S I A M N A S F F I G F F F S 104

401 TTATATAGCACTACTATTTCTTTATCCTTACTTACTATTAGCACTTTTCTTGGTTTAAAAAATCAACATGAGCTTACTATTCTGTTTACTTTTTTA 500  
AATATATCGTATGATAAAGAAAAATAGCAATCAATCAATCGTCAAAAGAACCAATTTTATGTTTGTACTCGAATCAATAAGACAAATGAAAAAT  
105 Y I A Y Y F F L S L L T I S S F S W F K X S M M S L V F L F T F L 137

501 TTTGTAGAATCCTTATTCTGGATTATCACTTGGACAAATCGGATAATTGGATTATTGCCAATTTTTCAGTATATGTTAAATCCAAATCCGTATGCAATTGA 600  
AAACATCTTAGGAATAAGACCTAAATAGTCAACTGTTTACCTTATTAACCTAATAACGGTTAAAAAGTCATATACCAATTAAGGTTAGGCAATGCTAACT  
138 F V E S L F M I Y O L D N G I I G L L P I F O Y M V M S M P Y A L I 171

601 TTTATTGGCTTACATTAATCTATCATAATCCATTGACTGTATTTCTGTTATAGAACTGGAGGAGAGTGTAAAAAGTTGGAATGGGAAAGTTAAG 700  
AAATAACCGAATGTAATGATAGATAGTATTAGGTAACGACATAAAGACAAGTATCTTGACCTCTCTCAGATTTTCAACCTTTACCCCTTTCAATTC  
172 Y M L T L L S I : I P L T V F S V M R N M R R V . 196

[illegible]

- 25 -

1301 AAGCAATAGATGACAAAGGTTCTTTTCAAGAAAAATACTATCCAGCTGTAAAAGAAAAAGGTTTATCGAACTGTTTGGCCAAAGCATTGACAGTTGCT 1400  
TTCTTATCTTACTGTTTCCAAAGAAAGCTTTTATGATAGGTGACATTTCCTTCCAAATAGCTTGAGCAAGCGGTTCCCTAACTGTCAAGCA  
414 C I E \* 417

gcp61

- 26 -

(SEQ ID NO: 65) 1 CTTCTTCCACCATTCACCAAGTCCTAGCACAGAAAAGAGTCTCTATCTTCAAAAGAAATTTATTACCTTTCCACATCTGACTTTGGTATTTT 100  
(SEQ ID NO: 66) CAAAAAAGCTGTAAGATTTTCAGCAATCTGTCTCTTTCTTCAGCAGATATCAAGCTTCTTTAAATAATGAAAGTGTAGACTGAAACCATATAAATAA

101 TAGAGAAAAATTAAGTTCTCCCATGTTTATGAGAGGTTCTCTTTATGCGAATGAAGATTAGTAGTGAATCTGGGAAATTCAGCTCCCAAAACAACT 300  
ATCTCTTTTAAATTCAGAGGTTACCAAAATACCTCTCAAGGACAAATACGCTTACTTCTAAATCATCACCTTAGACCTCTTAACTGAGCGGTTTGTTCCTA  
(SEQ ID NO: 64) 1 N V Y O E V P V Y A N E D L V V E S G K L T P K T S 36

201 TTTCAAATAACCGAGTGGCGCTTAAATAAACAAGGAATTCAGTATTTAAGCTATCAAAATCATCAATTTATAGCTGCGGACAAACGATTTTATATGATC 300  
AAAGTTTATTTGGCTCACCGGAAATTTATTTGCTTAAAGTCAATAATTCAGTAGTTAGTAAATATCAAGCGCTTGTTCCTAAAAATATACTAG  
27 F Q I T E W R L N K O G I P V F K L S N H O F I A A O R R F L Y D Q 60

301 AATCAGAGGTAAGTCCAACTAAATAAAGATATGTTAGAAATCTCACTTTAAACTGTACAAATAGTCTTATGATTTAAAGAGAGTGAATCATCTTATC 400  
TTAGTCTCAATGAGGTTGTTATTTTTCATACCAATCTTAGAGCAAAATTTGACATGTTTATCAGGAATAGTAAATTTCTTCACTTTAGTAGGAATAG  
61 S E V T P T I K R V M L E S D F R L Y N S P Y D L E V K S S L S 93

401 AGCTTATTCGCAAGTATCAATGCAACAGCACTTTTUTAGAGAGAGAGAAATTTCTACATATTCATCAGGCTGGATGGGTAAGTAAAGAAATCAACTTCT 500  
TCAATTAAGCTTATAGTATAGCTGTCTGTTACAAACATCTCTCTCTCTTAAAGATGTTATAGTAACTGCGACCTACCCATCGATTTCTTAGTTCAAGA  
94 A Y S O V S I D E T N F V E G R E F L H I D O A O W V A K E S T S 126

501 GAGAGAGATAATCGATGAGTAAAGTTCAAGAAATGTTATCTCAAAATAATCAGAAAGATCTTTCTCTATTTATGTTAAAGCAACTGACTACTGCGAAAG 600  
TCTCTTCTATTAGCTACTCTTCAAGTCTTTACAAATAGACTTTTATAGTCTTTCTAAGAAAGAGATAAATAGCAATTCGTTGAGTATGACTCTTTTC  
127 E E D N H N S K V Q E M L S E K Y O K D S F S I Y V R O L T T G K E 160

601 AAGCTGCTATCAATCAAGATGAAAGATGTTATGCGAGCGGCTTTTGAACCTCTCTTATCTCTATTTATCGGCAAGGAAATTAATGAGCGGCTTTATCA 700  
TTCAACCATAGTTAGTTACTTCTACTTTCTACATACCTCGGTCCGAAACCTTTGAGAGATAGAGATAATATCGGTTCTTTTATTTACTCCCGAAATAGT  
161 A G I N O D E R M Y A A S V L K L S Y L Y T O E K I N E G L Y O 193

701 GTTAGATACGACTGTAATAATCTGATCTGAGTCAATGATTTTCCAGGTTCTTATAAACAGAGGGAAGTGTAGTCTTCTTAAAAAGAGATATAATAA 800  
CAATCTATGCTGACATTTTATGCTAGACCTCACTTACTAAAGGTCAGAAATATTGCTCTCCCTCACCATCAGAGGATTTTCTTCTTATTTATTT  
194 L D T T V K Y V S A V N D F P G S Y K P E G S G S L P K K E D N K 226

801 GAATATTTTAAAGGATTTAATACCAAGATATCAAAAGAAATCTGATAATGAGTCTATAATCTATTGGGATATTACATTTCAAACCAATCTGATGCCA 900  
CTTATAAGAAATTTCTTAATTAATGCTTTCTAGTCTTTCTAGACTATTACATCGAGTATTAGATAACCTATAATGTAAGATTTGTTAGACTACGCT  
227 E Y S L K D L I T K V S K E S D N V A N N L L G Y Y I S N O S D A T 260

901 CATTCAAATCCAGATGCTCTGCCATTATGCGAGATGATTGGGATCCAAAGAAAAATTCGATTTCTTCTAAGATGGCGGGAAGTTTATGGAAGCTATTTA 1000  
GTAAGTTTAGGTTCTACAGCGGTAATACCTCTACTAACCTAGGTTTCTTTTAACTAAAGAGATTTACCGGCGCTTCAAATACCTTGGATAAAT  
261 F R S R M S A I M G D D W D P K E K L I S S K N A G K F M E A I Y 293

1001 TAATCAAAATGCAATTTCTGCTAGACTCTTCACTAAAAACAGATTTGATAGTCAGCGAATTCGCAAGGTTTCTGTTTAAAGTACCTCATAAATTCGA 1100  
ATTAGTTTACCTAAACAGGATCTCAGAACTGATTTTGTCTAAACATATCAGTCCCTTAAAGGTTTCCAGAAAGCAATTTCACTAGTATTTTAACTT  
294 N C K G F V L E S L T K T D F D S O R I A K C V S V K V A N K I G 126

1101 GATGCGATGAATTAAGCATGATACGCGTGTCTGCTATGCGAGTCTCTCATTTATCTTTTCTACTTAAAGAAATCTGATTATGATACGATTTCTA 1200  
CTAGGCTACTTAAATTCGTAATGCTATGCTTACCAACAGATACCTTAAAGGTTAAATAGAAAGATAAAGTGTATTCTTAAAGCTAATACTATGCTAAAGAT  
327 G A D E F K N D T G V V Y A D S P F I L S I P T R N S D Y D T I S K 160

1201 ACATAGCCAGGATGTTTATGAGGTTCTAAATGAGCGAACCAATTTTAAATCATTTCTCAAGAAAGGATATTTCAAAAAGCATGCTAAAGCGGCTT 1300  
CTATCGGCTTCAAAATACTCCAAGATTTACTCCCTTGGTCTAAAAATTTAGTAAAGAGTCTTCCCTATAAAGTTTTCGTAGGATTCGCGCAA  
361 : A R D V Y E V L R . 371

seq76

- 27 -

(SEQ ID NO: 68) 1 TTGAAAAATATTATCTATAAGAAACACATATAAATGTAAACAAAGCGTAAATATTTATTAGCGCTTTTCTTGTATACTAGTATTGTCTTTAAAGAAAGCA 100  
(SEQ ID NO: 69) AACTTTTATAAAGATATTCTTCTGTATATTTCATTTCTTCCGCAATTATAAATAATCCGAAAAAACCATATGATCATAACAGAAATTTTCTCTCT

101 GTATCTACCTAAATATGAAGAAAAAATCTTAGCTCAGCTTTATTAAGTACAGTAATGTTTCTCAAGTACCTGTTTAACTACCTCCGATCCAGAAACG 200  
CATAGATGCATTATCTTTCTTTTACAAATCCGAGTGAAATAATTCATGTTTACCAAGAGCTTCATCGACAAATTTGTACCGGTACGCTCTTTCTC

(SEQ ID NO: 67) 1 M K K K I L A S L L L S T V N V S O V A V L T T A N A E T 29

201 ACTGATGACAAATTTGCTCTCAAGATAATAAATTAGTAACTTAACAGCACAACAACAGAACCCAAAAACAAGTTTACCAAAATTCAGGACCAAGTAT 300  
TGACTACTGTTTAAACGACGAGTCTATTATTTTAACTCAATGAATTTGCTGTGTTGTTGTTCTTCCGCTTTTGTTCACCTGCTTTAAGTCTCTCTCATA

30 T D D K I A A Q D N K I S N L T A Q Q Q E A Q R O V D O I Q E Q V S 63

301 CAGCTATTCAAGCTGAGCAGTCTAACTTCAAGCTGAAATGATAGATTACAAACAGAAATCTAAGAAACTCCAGCGTGAGATTACAGAACTTTCTAAAAA 400  
GTGATAAGTTTCCGACTGCTCAGATTGAAGCTTCCGACTTTACTATCTAATGTTCTCTTAGATTCTTTGAGCTCCCACTTAATGTTCTTGAAGATTTTT

64 A I Q A E O S N L Q A E N D R L Q A E S K K L E O E I T E L S E N 96

401 CATTTGTTCTGTAACCAATCGTTGCAAAAACAGCTCGTAGTCTCAACAAATGGAGCGTAACCTAGCTATATCAATACCAATTTAACTCAAAATCA 500  
GTAAACAAAGACATTGCTTAGCAACCTTTGTTCCGAGCATCAGAGTTTGTTTACCTCGGCATTGATCGATATAGTTATGTTAAGCTTTGAGTTTATG

97 I V S R N O S L E K Q A R S A Q T N G A V T S Y I N T I V N S E S 129

501 ATTACAGAGCTATTTACGCTGTTGCTCAATCAGTGAATGTTATCTCCAAACAAACAAATGTTAGAACAAACAAAGGCAGATAAAAAAGCTATTTCTG 600  
TAATGTTCTTGATAAAGTCCAGAACGAGCTTACTCACTTAGCATAGAGCTTTGTTGTTTACAACTCTGTTGTTTTCCTCTATTTTTCGATTAAGAC

130 : T E A I S R V A A N S E I V S A N N E M L E Q Q R A D K K A I S E 163

601 AAAACAAAGTAGCAAAATATGATGCTATCAATCTGTAATGCTAATCAACAAAAATTTGCTGATGATGCTCAAGCATGCTACGAAACAGGCAGAACT 700  
TTTTGTTCTATCTGTTTATTACTACCATAGTATGACATTAACGATTAGTTGTTTAAACCGACTACTACGAGTTGTTAAGTATGCTTTGTCGTTCTTGA

164 K O V A N N D A I N T V I A N Q Q K L A D D A Q A L T T K Q A E L 196

701 AAAAGCTGCTGAATTAAGTCTTCTGCTCAGAAAGCAGTACCTCAAGCGGAAAAAGCAAGGCTATTAGAGCAGAAGCAGCAGCTCAGGCAGAGGCTCG 800  
TTTTCCAGGACTTAATTCAGAACGAGCTCTTTGCTGATCGACTTCCCTTTTCTTCCGATAATCTGTTCTTCTGCTGAGTCCGCTCTCCAGC

197 K A A E L S L A A E K A T S . 211

- 28 -

YNES\_BACIU

(SEQ ID NO: 71) 1 ATGTTAATTCCTTTATTGATTATTTGGCTACTTCATAGGAGCATTCCATCTGCTTAATTGTGGGCAAGCTTCCAAAGGAATTGATTTGGGAGC 100  
(SEQ ID NO: 72) TACAAATTAACGAAATAACTAATAAAACCGATGAATATCGTGTAAAGTAGACCGAATTAACACCCGTTTGGAAACGTTTCTTAACATAAGCCCTCG 100  
(SEQ ID NO: 70) 1 M L I A L L I I L A Y L I O S I P S Q L I V O K L A R G I D I R E N 16  
101 ACGAAGCCCAACTTAAGGCTACCAATGCATTCCGTACATTTGGGTGTAAGAGCTGGTTGGTCTGCTATAGCCCGAGATATTTTGAAGGCACTGGC 200  
TGCCTTCCCGCTTCAATCCCGATGTTTAAAGCATGTAACCCACATTTTGGACCAAGCCAGCAGTATCGGCTCTATAAAACTTTCCCTGTGACCG  
35 G S N L G A T N A F E T L S V K A G S V V I A G D I L E G T L A 67  
201 AACTGCATTGCTTTTCTCATGCATGTTGATATTACCCGCTTCTTCCAGGAGTCTTTGGGTTTAAAGGCACTGTTTCCCATCTTCCGAAATTTAAA 300  
TTGAGGTAAAGGAAAGAGTACGTACAACTATAAGTGGGCAAGACGTTCTCAGAAACCCCAAAATCCGTTGCACAAAGGTAGAGCGGTTTAAATTT  
68 T A L P P L M N V D I N P L L A G V P A V L G H V P P I F A K F K 100  
301 GCGGTAAAGCGGTTGGCAGATCAGGAGCGCTTTTCTATTTTACGACCCCTGTTTATTTATCAGGATGGTTGGGTTATTTCTTCATCTTTTATACTTGA 400  
CGCCCATTTCCGACCGCTGTATGTTCTCCGCAAAAGATAAATGGTGGGCAATAAATAGTGTCTACCAACCCCTATAGAGGTAGAAAAATATGAACT  
101 G G E A V A T S G G V L L F Y A P L L F I T N V A V F P I F L Y L T 134  
401 CTAAATTTGTTTCTCTCATGGATGTTTAAAGGATCTATACCTGTTATATATAGTTTCTTTGTCATGATAGTATTATTGATTTGTTACCTGCT 500  
GATTTAAACAAAGAGAGAGTAGCTACAAATTTCCCTAGATATGCAATATATATCAAGAAACAGGTACTATCCATAAATAACTAACAGCAATGGGACGA  
135 K F V S L S S M L T G I Y T V I Y S F F V H D T Y L L I V V T L L 167  
501 CACTATTTTGTGATATACAGACCGAGCAACATTAAAGCAATTATCAATAAAACAGAACCTAAAGTAAATGTTTATAA 582  
GTGATAAAACACTATATGTTCTGTGGCTGCTTGTAAATTTGCTTAATAGTTATTTTGTCTTGGATTTTCATTTTACCAATATT  
168 T I F V I Y R N R A M I R I I N K T E P K V K W L . 193